REPORT RESUMES

ED 002 948 24

RESEARCH ON THE CHARACTERISTICS OF EFFECTIVE COLLEGE TEACHING.

BY- MCKEACHIE, WILBERT J. AND OTHERS

MICHIGAN UNIV., ANN ARBOR

REPORT NUMBER CRP-850 PUB DATE 64

EDRS PRICE MF-$0.45 HC-$11.88 297P.

DESCRIPTORS- *EFFECTIVE TEACHING, TEACHER EVALUATION, *TEACHER CHARACTERISTICS, STUDENT CHARACTERISTICS, *COLLEGE INSTRUCTION, LEARNING PROCESSES, MEASUREMENT INSTRUMENTS, *COGNITIVE MEASUREMENT, COURSE OBJECTIVES, PSYCHOLOGICAL PATTERNS, TEACHING METHODS, *STUDENT TEACHER RELATIONSHIP, PERSONALITY ASSESSMENT, ACADEMIC ACHIEVEMENT, PREDICTIVE MEASUREMENT, ANN ARBOR, MICHIGAN

THE EFFECTS OF DIFFERING TEACHER CHARACTERISTICS AND OF VARIOUS STUDENT CHARACTERISTICS UPON LEARNING IN COLLEGE WERE MEASURED BY AN INSTRUMENT CALLED THE "CRITERIA TEST," WHICH CONTAINED SUBTESTS FOR MEASURING SEVERAL LEVELS OF COGNITIVE COURSE OUTCOMES AS FUNCTIONS OF INTERACTIONS AMONG METHODS OF TEACHING, TEACHER PERSONALITY, AND STUDENT PERSONALITY. SIXTEEN OBJECTIVES OF AN UNDERGRADUATE, GENERAL PSYCHOLOGY COURSE WERE TO BE USED IN THIS EVALUATION. ONLY NINE OF THESE, HOWEVER, PROVED TO BE OF SIGNIFICANT IMPORTANCE AFTER OPINIONS WERE GATHERED FROM PSYCHOLOGY INSTRUCTORS THROUGHOUT THE UNITED STATES. AVAILABLE DATA WERE ALSO COLLECTED ON SOME SPECIFIC TEACHING METHODS AND THE ACHIEVEMENT MOTIVATION, AFFILIATION MOTIVATION, AND ANXIETY OF STUDENTS WHO WERE EXPOSED TO THEM. IN ADDITION, PERSONALITY MEASURES WERE ADMINISTERED TO TEACHERS AND STUDENTS, AND OBSERVATIONS WERE MADE OF STUDENTS' AND TEACHERS' RESPONSES IN THE CLASSROOM. ALL OF THESE FACTORS WERE USED IN THE MAKEUP AND ADMINISTRATION OF THE "CRITERIA TEST." AFTER EVALUATING ALL TEST INFORMATION THE INVESTIGATOR FOUND NO SINGLE VARIABLE OR COMBINATION OF VARIABLES TO BE AS POWERFUL AS INTELLECTUAL ABILITY IN ACCOUNTING FOR STUDENT ACHIEVEMENT IN COLLEGE. STUDENT RATINGS OF TEACHER CHARACTERISTICS, TOO, WERE SUGGESTED AS EXCELLENT MEASURES FOR TEACHER EFFECTIVENESS AND, THUS, STUDENT PERFORMANCE. PERSONALITY COMPATIBILITIES WERE ALSO SIGNIFICANTLY RELATED TO POSITIVE STUDENT ATTITUDES AND ACHIEVEMENT. (JH)

ED002948

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

RESEARCH ON THE CHARACTERISTICS OF EFFECTIVE COLLEGE TEACHING

Cooperative Research Project No. OE 850

Wilbert J. McKeachie
Robert L. Isaacson
John E. Milholland

The University of Michigan
Ann Arbor, Michigan

1964

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.

Contents

(Pages within sections are separately numbered.
The section number appears on each page.)

Section	Title	Authors
I.	Introduction	
I-A	Research on Teaching at the College and University Level. Reprinted from: N. L. Gage (ed.) *Handbook on teaching* Chicago: Rand McNally, 1963, pp. 1118-1172	W.J. McKeachie
I-B	Motivation, Teaching Methods, and College Learning. Reprinted from M. R. Jones (ed.) *Nebraska symposium on motivation:* 1961 Lincoln: Univer. Nebraska Press, 1961. pp. 111-146.	W.J. McKeachie
I-C	The Michigan Project: Research on Characteristics of Effective College Teaching.	
II.	Outcome Measures	
II-A	Measuring Cognitive Abilities	J.E. Milholland
II-B	Attitudinal Studies	Dodd H. Bogart
III.	Chronology and Description of Samples Used	
IV.	Teacher Characteristics Related to Effectiveness	
IV-A	The Dimensions of Student Evaluations of Teaching	R.L. Isaacson W.J. McKeachie J.E. Milholland Yi-Guang Lin Margaret Hofeller James W. Baerwaldt Karl L. Zinn
IV-B	Correlation of Teacher Personality Variables and Student Ratings, Reprinted from J. Educ. Psychol. 1963, 54: 110-117	R. L. Isaacson W. J. McKeachie J. E. Milholland
IV-C	Student Evaluations of Recitation Sections in Beginning Economics	Robert L. Isaacson

V. Student Characteristics Related to Achievement and Curricular Choices	
V-A Student Characteristics Related to Achievement (Main Effects)	Yi-Guang Lin
V-B Relations Between n. Achievement, n. Affiliation, Debilitating Anxiety, and Several Measures of Collegiate Performance	Robert L. Isaacson W.J. McKeachie John E. Milholland Yi-Guang Lin
V-C A Model for Curricular Choices	Robert L. Isaacson
V-D The Relation Between n. Achievement, Test Anxiety, and Curricular Choices	Robert L. Isaacson
VI. Interactions of Student and Teacher Characteristics with Outcome Measures	
VI-A Teacher Characteristics and Student Measures from the TAT	
VI-A-1 Student Affiliation Motives, Teacher Warmth, and Academic Achievement	W.J. McKeachie
VI-A-2 Student Achievement Motives, Achievement Cues, and Academic Achievement	W.J. McKeachie Robert L. Isaacson John E. Milholland Yi-Guang Lin
VI-A-3 Student Power Motives, Discussion Teaching, and Academic Achievement	W.J. McKeachie
VI-B Interactions of Teacher Characteristics and Students' Scores on the Opinion, Attitude and Interest Survey with Outcome Measures	John E. Milholland
VI-C Student Ratings of Teachers and Student Performance as a Function of Student and Teacher Personality Factors	James W. Baerwaldt John E. Milholland W. J. McKeachie
VII. The Matching of Students and Teachers so as to Enhance Student Achievement	John E. Milholland

Research on Teaching at the College and University Level[1]

W. J. McKEACHIE
University of Michigan

The ultimate criteria of effective college teaching are changes in students toward the goals of higher education. Research on college teaching begins with consideration of institutional objectives and course goals. While it is not appropriate to discuss the goals of higher education here, it is important to note that they include attitudinal and emotional changes as well as more obvious cognitive goals such as critical thinking and broad knowledge. In evaluating the effectiveness of college instruction we need to consider not only the accumulation of knowledge but the development of problem-solving skills and desirable attitudes.

Because different professors weight different types of goals differently, some see the professor primarily as an information-giver, others as a group leader, and others as a nondirective psychotherapist. Each endeavors to fulfill the particular role he sees as most important in effecting his students' education.

But teachers seldom ask themselves "What do I contribute to the educational process?" They make assignments, lead discussions, reflect student comments, give and correct tests—all in the faith that these activities contribute to education. Is their faith justified? Is each activity important and worth while? Does it matter which methods they use?

This chapter reviews the evidence bearing on these questions. General principles of learning are relevant to our use of teaching methods, but there is little empirical evidence to guide one's choice of method.

LEARNING PRINCIPLES RELEVANT TO TEACHING METHODS[2]

Let us first review briefly some principles of learning believed to be important in effecting the student's achievement of the objectives of education. What principles apply to how a student acquires a knowledge and understanding of chemistry? An appreciation of music? We shall consider motivation, organization, variability, verbalization, feedback, contiguity, and active learning.

---

[1] The author gratefully acknowledges the able assistance of Mrs. Virginia Lickey in compiling the bibliography upon which this chapter is based. Portions of the literature survey were carried out in connection with projects supported by the Lilly Endowment, the Fund for the Advancement of Education, and the U. S. Office of Education. Major portions of this chapter are also contained in the author's book, *Teaching Tips* (4th ed.; Ann Arbor, Mich.: Wahr Publishing Company, 1960), and in the author's chapter in Nevitt Sanford (Ed.), *The American College* (New York: Wiley, 1962).

[2] Portions of this section are based upon W. J. McKeachie. *How do students learn?* in R. Cooper (Ed.), *The Two Ends of the Log* (Minneapolis: Univer. of Minnesota Press, 1958), pp. 26–39.

Motivation

We know that student learning and memory are closely tied to ***motivation***. Students usually learn what they want to learn, but they often have great difficulty learning material which does not interest them. Most of us have to recognize that not all students are deeply interested in everything we want to teach them.

One primary problem, then, is motivating students. Usually the learning psychologist stops with this point, but to be useful the principle of motivation needs to be accompanied by information about dependable motives of college students. Most students are taught by their parents to want to do well in school and hence have some motivation for school achievement.

Also, most students want to be liked. This motive may work against teachers as well as for them. The teacher's friendly approval can be an important reward for learning, but the "average-raiser" is not always well liked. Some students who seek acceptance by their classmates may avoid any conspicuous display of academic achievement. Many students suffer conflict between the need to get good grades and the need to be well liked. One symptom of the conflict is the ostentatious neglect of study by some bright students and their apparent surprise when they get good grades—a "ploy" carefully analyzed in Potter's ***One-upmanship*** (1955).

Many students have other conflicting motives. Conflict between desire for independence and dependence occurs often; students are likely to resent a teacher who directs their activities too closely, but they also are likely to be anxious when given too much independence. The teacher then has the problem of simultaneously satisfying both needs. As a result of this conflict, some students disagree with the teacher, not on any rational grounds but simply as a way of expressing emotions. Similarly, student apathy in a required course may be an irrational expression of resentment at being required to do anything.

Any thoughtful college teacher could list many other common student needs. The point is simply that these needs provide the tools by which teachers get learning to take place. As far as learning or retention is concerned, it matters little which motives are involved—the important thing is that motivation exists and that conflicts do not become activated.

Let us then consider the case of our most important motivational device—grades. Whatever a student's motivation for being in college, grades are important to him. If he is genuinely interested in learning, grades represent an expert's appraisal of his success; if he is interested in getting into professional school, good grades will unlock graduate school doors; if he wants to play football, grades are necessary to maintain eligibility. Most students are motivated to get passing grades, and thus grades are a powerful motivational tool for teachers.

Most college teachers are a little embarrassed by this. They regard grades as one of the necessary evils of teaching. They try to discount grades in discussing the organization of the course; they try to arrive at grades in such a way as to avoid trouble with disappointed students. But they also frequently fail to use grades to bring about the sort of learning they desire.

Because grades are important to them, students will learn whatever is necessary to get the grades they want. If teachers base their grades on memorization of details, students will memorize the text. If students believe grades are based upon their ability to integrate and apply principles, they will try to acquire such ability.

As far as speed of learning goes, it does not matter what motives are aroused. Yet this does not mean that the type of motivation used may not affect other aspects of learning. For example, a good deal of evidence has accumulated to suggest that negative and positive motives affect behavior differently. When negative motives predominate, students will work hard, but only if this is the only way to avoid undesirable consequences. If there are ways out of the situation, students will work hard, but only if this is the is that students do as little work as they can get away with and cannot develop a beneficial interest in the subject.

The closer the goal or threatened danger, the higher motivation is. But negative motives are not as effective outside the learning situation as are positive motives because the effectiveness of fear depends more upon the closeness of the goal than does the effectiveness of positive motives. Thus the teacher who motivates his students by fear of bad grades or of reprimands must use frequent tests if his threats are to be taken seriously by his students.

The striking difference between the behavior of a student motivated by fear and one motivated by hope is illustrated in their behavior during examinations. Atkinson and Litwin (1960) showed that male students who were high in anxiety about tests, as measured by the Mandler-Sarason Test Anxiety Questionnaire and the TAT, were among the first to complete the course examination and tended to do poorer work on the examination than they did during the course. Students with positive motivation to succeed tended to stay in the examination room longer. The fearful students tended to avoid the situation which aroused their anxiety.

Thus, motivation is important in learning. We can use student motivation for success and approval to produce learning, and grades can be important incentives for both negative and positive motivations. Thus it is important to make sure that grades are not separate from the kind of learning we want to occur. Using grades chiefly as a threat may produce avoidance rather than interest.

Organization

A teacher's job is not done when he arouses the interest of his class, for the amount students learn depends in part upon the amount he teaches. But this relationship is not so simple as it may at first appear. It may well be that the *more* we teach, the *less* our students learn.

In *The Great Didactic*, Comenius writes:

> If we take a jar with a narrow mouth (for to this we may compare a boy's intellect) and attempt to pour a quantity of water into it violently, instead of allowing it to trickle in drop by drop, what will be the result? Without doubt the greater part of the liquid will flow over the side, and ultimately the jar will contain less than if the operation had taken place gradually. Quite as foolish is the action of those who try to teach their pupils, not as much as they can assimilate, but as much as they themselves wish . . . (Comenius, trans. by M. W. Keatinge, 1931, pp. 87–88).

Some years ago, several teaching fellows were arguing about how to teach about the nervous system. One group argued that since students do not remember many details, teachers might better omit them and teach only the essentials which they wanted everyone to learn. Another group argued that students forget much of what they learn. "But," they said, "if they're going to forget a large percentage, we need to teach much more than we expect them to remember. Otherwise they'll forget even the important things."

Since such an argument invited an experiment, both groups agreed to try out their ideas in their own classes and compare the results on the final examination questions covering the nervous system. The outcome was clear: students whose instructor had omitted details were clearly superior to those whose instructor had given them the whole story. This result would not have surprised Katz (1950), the German-Swedish psychologist who devised a number of experiments demonstrating that beyond a certain point adding to the elements in an intellectual task causes confusion and inefficiency. Katz called this phenomenon "mental dazzle."

Fortunately, it is possible to teach more and have it remembered better. The key factor is "Organization." As Katona (1940) demonstrated, in a series of experiments on organization and memory, we can learn and remember much more when our learning fits into an organization. If one heard a series of numbers chosen at random, like 738105477, and was asked later what the fourth number was, he would probably have difficulty remembering it. But given the numbers 248163264, along with the principle of the series, he could remember immediately what the fourth number was. Teaching which helps students find a framework within which to fit new facts is likely to be much more effective than teaching which simply communicates masses of material in which the student can see no organization. The successful teacher is one who enables students to see meaningful problems. The ideal class would begin with a problem so meaningful that the students would always be just a step ahead of the teacher in approaching a solution.

Thus *motivation* and *organization* are two key concepts which should guide efforts to help students acquire knowledge.

Variability and Verbalization

Another group of objectives stems from the traditional intellectual objectives of education. How can we help students develop principles and concepts which they can apply to a wide range of problems? Many teachers have been disheartened by having a student answer a routine problem perfectly and then fail to use the same knowledge in solving a new problem to which it is also relevant.

A number of attempts to improve this kind of learning have been made. The theory of "learning by doing" held that if one learned something in the situation where the learning was to be used, he would not have the added step of learning when to apply it. At first glance, this seems perfectly reasonable, and it even makes sense psychologically. The only problem is that the number of situations in which one must use knowledge is infinite. Our whole civilization is based on the fact that man can use words to short-cut the long process of learning by trial and error. Direct experience may be useful at certain stages of learning. If we are to learn to apply a principle in new situations, we need to develop it from experiencing specific instances of the principle in *varying* contexts. A number of experiments have demonstrated that repetitive drill is much less effective than varying problems in developing principles which can be applied to new situations (e.g., Wolfle, 1935).

Verbalization can help the student identify the common elements in these situations and shorten the learning process. In fact, research suggests that even such a complex skill as flying an airplane can be learned in shorter air time if the learner practices the skill mentally and verbally (Flexman, Matheny, & Brown, 1950). While verbalizing principles is undoubtedly better than unresolved confusion, Kersh (1958) suggested that allowing students to discover principles for themselves may enhance motivation.

Feedback, Contiguity, and Active Learning

If students are to learn skills, they have to practice the skills, but practice alone does not make perfect. Practice works only if the learner *sees the results* of his practice, i.e., if he receives feedback (Woodworth & Schlosberg, 1954).

Feedback is most effective when it is contiguous to the response being learned. A major feature of teaching machines is that the learner finds out quickly whether his response is right or wrong.

A number of experiments have demonstrated that *active* learning is more efficient than passive learning (McGeoch & Irion, 1952). One reason for this may be that active learning provides improved opportunities for feedback. Discussion techniques may help develop critical thinking because students do the thinking and can check their thinking against that of others. But one danger of "student-centered" or "non-directive" discussion is that the *results* are not apparent. Students may make comments, express opinions, and participate actively, but their opinions may not be any more informed at the end of a semester than at the beginning. Of course not all feedback has to come from the instructor—students can learn much from other students or from books—but, in order to learn, students need to test their ideas in a situation in which they can get the results of the test.

Nevertheless, we need to go a step beyond the principle that students learn what they practice when they see the results of their practice. It is not always easy to get students to practice critical thinking in the classroom. The student who remains quiet in class avoids the risks of disagreement, criticism, and embarrassment. To develop critical thinking, the student must learn to want to think.

This brings us to another category of goals of education—developing interests, changing attitudes, creating motivation. There are many ways to do these things. To develop a student's interest in thinking, we have to make it satisfying. A smile, a nod of encouragement, an excited "Good! Let's follow that idea through"—these are the tools that teachers use, not only to provide knowledge of results, but to develop the motivation to continue intellectual activity.

To develop motivation, teachers must pose problems within the range of their students' abilities. Studies of the development of achievement motivation in children indicate that parents develop this motivation by encouraging the child to do well and setting standards which the child can achieve. Other parents orient their children toward achievement but are disappointed in their aspirations for their children because they set unreasonable goals. To motivate students first to develop the ability to think critically and then to engage in critical thinking, teachers must give them experience in solving problems within their ken. This by no means implies that the student should never experience failure or criticism, but it does mean that the problems which he faces should, more often than not, be soluble.

One misconception of educators who have recognized the importance of motivation is that they have stopped with the motives students already have. Teachers have assumed that students will learn only those things in which they are interested so that one need only turn students loose in an environment which contains possibilities for learning. But a student's motives are not fixed; teachers can create new motives. We can teach students to enjoy learning for its own sake. Thus, while we must make use of existing motives to create initial satisfactions in learning, our efforts need not be limited by them.

METHODOLOGICAL CAUTIONS

Before turning to the research evidence on the effectiveness of teaching methods, let us review a few methodological points. Determining which of two teaching methods is more effective looks like a simple problem. Presumably, all that is necessary is to teach something by one method and then to compare the results with those obtained by teaching the same thing by another method. This is essentially the research design of many studies widely quoted as demonstrating the effectiveness of television, discussion, independent study, or other methods. Unfortunately, there are pitfalls which enthusiasts for one method or another are likely to overlook.

Suppose, for example, that students are given an opportunity to take a class taught by some method quite unusual in their college. The very fact that the method is different gives it excitement. Sometimes students react to a new method with enthusiasm, at other times, with outraged hostility. A hostile reaction seems particularly likely when students taught by traditional methods. In any they are competing on examinations with students taught by traditional methods. In any case, it is difficult to know how much of student improvement (or loss) in learning may be accounted for by emotional reaction to a new and different method and how much we can expect when the new method is routine. This "Hawthorne effect" influences professors as well as students. How many new curricula, new courses, or new teaching methods have flowered briefly and then faded as the innovators' enthusiasm waned or as new staff members replaced the originators? Unfortunately, relatively few studies have made comparisons over a period longer than one semester. As we shall see later, students who have experienced a semester of instruction by a new method (except television) are generally more likely to choose a section taught by this method than are students without such previous experience. This difference in motivation of students as well as their added skill in a new method might result in even greater advantages for the new method after two or more semesters of trial than after a single semester.

A second problem is that of establishing a suitable control group. In some experiments a single instructor uses both teaching methods. Here it is obviously difficult for the experimenter to determine how much the instructor's own personality and skills influence the outcome. One remedy for this difficulty is to persuade several professors to use both methods. Aside from the salesmanship necessary to institute such a research design, the effort involved in trying to teach by two methods, keeping strictly to each, is tremendous. As a result, the methods either tend to coalesce or the experimenter institutes additional, artificial constraints to accentuate the differences between them.

Another problem in establishing controls may arise when the conditions of the experiment introduce special factors which interfere with normal results. For example, the experiment may require extensive testing, the presence of observers in the class, or other interferences with normal classroom routine. A class in which a "live" professor is talking to television cameras is probably not a suitable comparison group for classes watching the lesson on television receivers.

A fourth problem, biased sampling, needs only brief mention. Such a problem would arise in studies of educational television aimed at determining whether students taking the course at home learn as much as those taking it on campus. People who sign up for a television course and come to campus to take the examination are probably somewhat different in motivation and background from typical college sophomores. As Greenhill (1959) pointed out, efforts to equate such groups are rarely successful.

A fifth problem lies in the statistical methods used to analyze the results of teaching methods experiments. Ordinarily, we carefully try to avoid the type of error involved in concluding that one method is more effective than another when in reality they do not differ significantly. However, we are less likely to be sensitive to another type of error which may be just as damaging—the error of concluding that there is no difference in effectiveness when two methods are not found to differ significantly. In addition to the logical fallacy involved in accepting failure to disprove the null hypothesis as proof of no difference, there is the problem of choice of methods of analysis. The chance of obtaining such results depends upon the type of statistical analysis one uses. If one uses "weak" statistics, a difference is less likely to be detected than if one uses "strong" statistics. The true effect of a variable may be clouded if no effort is made to remove other sources of variance. As is suggested later, multivariate statistics (such as analysis of covariance) might, by taking out other sources of variance, reveal more clearly the true effects of varying methods. Further, when several tests of the same hypothesis are made with different groups, experimenters might well use combined tests of significance. For example, if 10 groups come out in the same direction, it is extremely unlikely that the methods do not differ in effectiveness even though no one difference would be statistically significant alone.

Sixth is the problem of dealing with interactions among teaching methods, student characteristics, teacher characteristics, or other variables. As we shall see in the section on student characteristics, a method which is effective for the learning of some students may be ineffective for others in the same classroom. Ordinary group comparisons between teaching methods may thus hide very important differences in the effects of the methods upon a particular type of student.

The Criterion Problem

The major problem in experimental comparisons of teaching methods is the criterion problem. Stuit and Wilson (1946) found that prediction of success in naval training improved as the criterion became increasingly well defined. Undoubtedly, one reason for the many nonsignificant differences in studies of teaching is poor criterion measures. The criterion problem is further illustrated by the experiment of Parsons, Ketcham, and Beach (1958). To determine the effectiveness of various methods, they set up groups in which students did not come to classes at all. The groups who did not come to class did *best of all* on the final examination. The examination was based entirely upon the textbook, however. As Parsons and Ketcham pointed out, their results with the other groups suggest that, as more and more new ideas and points of view are introduced, students become less and less likely to remember what the textbook says. This points to the problem of evaluation of effectiveness. If our goal is to have students remember the textbook, a test on the textbook is appropriate, but we cannot conclude that a particular method is superior in achieving all goals if we have measured only one.

Few professors complain that students are too highly motivated. Yet for purposes of research the degree of student motivation for good grades may actually make it very difficult to evaluate the effectiveness of two teaching procedures. Because passing or excellent grades are so important to students, they may compensate for ineffective teaching by additional study in order to pass the course examination at the level to which they aspire. Thus the effects of ineffective procedures may be masked or even misinterpreted when course examinations are used as criterion measures. Nachman and Opochinsky (1958) provided a neat demonstration of this when they found differences between a small and large class on surprise quizzes but no difference on a final examination. When significant differences in achievement are found in an experiment, the difference may simply reflect the degree to which students in differing classes were able to find out the content of the examination and the degree to which it would determine their course grade.

The difficulty in arriving at an over-all index of teaching effectiveness is complicated by the probability that a teacher who is effective in achieving one course objective is not necessarily effective in achieving others. Bendig (1955), for example, found a significant interaction between instructors and tests in an introductory psychology course. Some instructors' students did particularly well on certain tests during the course but not well on other tests. Cross (1958) and McKeachie (1959) found that instructors whose students did well on an objective test in psychology were ineffective when their students' achievement was measured on an essay test designed to measure understanding and integration of the materials. Thus, in studies of teaching it is important to specify objectives and to use measures of each objective. Measures of retention after the end of a course would often add to one's confidence in reported differences.

Because achievement measures have been so insensitive to differences in teaching methods, many experimenters stress the favorable student reactions to the new method they have introduced. Although the relationship between student satisfaction and learning is low (Elliott, 1950), it certainly can be argued that, assuming equal learning between two methods, we would prefer to have students leave our classes with warm feelings about their experience. Moreover, we would expect this feeling to be related to interest in learning more, and there is some evidence to support this expectation (McKeachie & Solomon, 1958). When, however, we use student satisfaction as a criterion, we should be aware that it is highly influenced by students' role expectations of college teachers. Marked deviations from these expectations are almost inevitably rated lower than more conventional teaching behavior. Laboratory studies of problem-solving groups reveal that group members rate authoritarian leaders as more efficient than democratic leaders (Haythorn, et al., 1956). This makes sense because of our expectations for leaders and also because a leader who plays an active role almost inevitably makes a more vivid impression on a group than a leader whose behavior is more subtle. In evaluating student reactions, therefore, one needs to be conscious of these role expectations and to determine a proper base line against which to evaluate the reactions.

As an aside here, let us also note that new methods are not usually tested except by a teacher who is enthusiastic about them. Consequently, we may be comparing student reactions to a new method and an enthusiastic teacher with reactions to an old method taught unenthusiastically.

The prospective researcher also needs to be warned that even a careful definition of desirable outcomes does not solve the criterion problem. In many cases, laudable attempts to measure attitudinal or affective outcomes have led to the conclusion that neither of two teaching methods was superior in achieving this or that goal, when there is no evidence that *any* teaching could affect achievement of the goal as measured by the tests used. At the very least, the experimenter needs to report some index of reliability; even better would be some evidence that the measure is at least sufficiently sensitive to reveal significant changes which occur between the beginning and end of the semester. If there is no change on a variable over a semester, it is unlikely that two teaching methods will differ in the amount of change they yield.

Finally, let us remind ourselves that evaluation need not end with tests given to the students who are enrolled in the experimental classes. In a large university it is easy to assume that an experimental course is assimilated into the whirlpool of teaching activity without even a ripple. Seldom, however, has this assumption been tested, and in smaller colleges or with large-scale innovations, it is not a safe assumption. We might gain much useful knowledge by looking outside our experimental classrooms to other effects of the experiment. Do students taught by one method rather than another make more use of their knowledge and skills in other courses they are electing? Is superior achievement in the experimental course won at the expense of achievement in other courses? What is the impact of a particular teaching method upon other faculty members? How does the use of a new method like television change faculty perceptions of teaching and its value? How does it affect faculty-administration relationships? In short, what effects does a new method have upon the total culture of the college?[3]

With these caveats in mind, let us review some of the research on college teaching.

RESEARCH ON TEACHING METHODS

This review is organized in terms of traditional categories of teaching method—lecturing, discussion, laboratory—and also in terms of fairly recent approaches to college instruction—project methods, independent study, and automated techniques.

LECTURING

College teaching and lecturing have been so long associated that when one pictures a college professor in a classroom, he almost inevitably pictures him as lecturing. The popularity of the lecture method probably derives from a conception of the instructor's primary goal as that of transmitting knowledge.

---

[3] Morris Janowitz started my trains of thought along this line.

Since students typically have few opportunities to make responses during lectures, they seldom receive feedback from their efforts to learn during lectures. Delay of feedback may not, however, seriously hinder the learner in acquiring knowledge if he is motivated and the material is not too difficult. We would expect lack of feedback to be a greater handicap if the lecturer's goal is to develop concepts or problem-solving skills. There is experimental evidence that, when this is the goal, active participation on the part of the learner is more effective than passive listening or observing.

Lecture Method versus Discussion Method

The effectiveness of lecture and discussion methods has often been compared. Since discussion offers the opportunity for a good deal of student activity and feedback, it could, in theory, be more effective than the lecture method in developing concepts and problem-solving skills. However, since the rate of transmission of information is slow in discussion classes, we would expect lecture classes to be superior in helping students acquire knowledge of information.

Although many studies have been made comparing the lecture method with discussion or other methods, few have used independent measures of these different types of outcome. The results of these studies are generally in line with our hypotheses but are certainly not conclusive. For example, using tests of information, Remmers (1933) found slight but nonsignificant differences favoring learning in large lecture groups as compared with that in small (35–40) recitation sections. Spence (1928) obtained similar results comparing lecture and discussion technique in classes of over a hundred students. In one of the earliest comparisons of lecture and discussion methods, Bane (1925) found little difference between the methods on measures of immediate recall but a significant superiority for discussion on a measure of delayed recall. Ruja (1954), however, found that the lecture was superior to discussion as measured by a test of subject-matter mastery in a general psychology course. In the other two courses in his experiment there were no significant differences in achievement, nor were there differences in changes in adjustment in any of the courses. Eglash (1954) found no difference between a discussion class and lecture class in scores on the final examination, in scores on an achievement test administered several weeks after the course had ended, or in scores on a measure of tolerance. Husband (1951) also found no significant difference in achievement of students in large (200-student) lecture and in "small" (50-student) recitation classes, but in five out of six semesters the lecture group was nonsignificantly superior. In all these experiments, the information measured by the examination could be obtained from a textbook, and in only one was a discussion group of less than 35 used.

When we turn to measures of more complex outcomes, the results are somewhat different. Hirschman (1952), using a measure of concept learning, compared the effectiveness of presenting material by dictation with that of presenting written materials followed by discussion and rereading. The reading-discussion method resulted in superior ability to identify examples of the concepts presented. In quite a different type of experiment, Barnard (1942) compared the effectiveness of a lecture-demonstration teaching method with that of a problem-solving developmental discussion in a college science course. In this experiment the lecture-demonstration method proved superior on a test of specific information, but the discussion method proved to be superior on measures of problem solving and scientific attitude. Similarly Dawson (1956) found problem-solving recitation and lecture-demonstration methods to be equally effective in a course in elementary soil science as measured by a test of recall of specific information, but the problem-solving method was significantly superior as measured by tests of problem-solving abilities. Other evidence favoring discussion was the experiment of Elliott, who found that students in his discussion groups in elementary psychology became more interested in electing additional courses in psychology than did students in a large lecture.[4] The results of DiVesta's (1954) study of a human relations course tended to favor a discussion method over the lecture method in improving scores on a leadership test. Similarly, Casey and Weaver (1956) found no differences in knowledge of content but superiority in attitudes (as measured by the Minnesota Teacher Attitude Inventory) for small-group discussions as compared to lectures.

Despite the many findings of no significant differences in effectiveness between lecture and discussion, those studies which have found differences make surprisingly good sense. In only two studies was one method superior to the other on a measure of knowledge of subject matter; both studies favored the lecture method. In all six experiments finding significant differences favoring discussion over lecture, the measures were other than final examinations testing knowledge.

When one is asked whether lecture is better than discussion, the appropriate counter would seem to be, "For what goals?"

Distribution of Lecture and Discussion Time

Many universities and large colleges divide class meetings between lectures and discussions. This administrative arrangement is supported by a study in the teaching of physiology, in which discussion meetings were substituted for one-third of the lectures (Lifson, Rempel, & Johnson, 1956). Although there were no significant differences in achievement, as compared with all lectures, the partial discussion method resulted in more favorable student attitudes which persisted in a follow-up study two years later. Warren (1954) compared the effectiveness of one lecture and four recitations with that of two lectures and three demonstrations per week. In one out of five comparisons, the one-lecture plan was superior, while the other comparisons found nonsignificant differences. Superior students tended to prefer the two-lecture plan while poorer students did not. On the other hand, in Remmers' (1933) comparison of two lectures and one recitation versus three recitations, the poorer students tended to do better in the lecture-recitation combination, although students in both groups preferred the three-recitation arrangement. In Klapper's study (1958) at New York University, most students preferred a combination lecture-discussion method to one employing all lectures or all discussions. Students at the State University of Iowa also preferred all group discussion or a combination of lecture and discussion to lectures alone (Becker, Murray, & Bechtoldt, 1958).

In a course in which the instructors wish not only to give information but also to develop concepts, the use of both lectures and discussions would thus seem to be a logical and popular choice. The lecture can effectively present new research findings; the discussion can give students opportunities to analyze the studies, find relationships, and develop generalizations. By participating actively in discussion, the students should not only learn the generalizations but should also begin developing skill in critical thinking.

The Lecturer versus Automation

Since we have hypothesized that the lecturer is particularly effective in transmitting information, it is in order to ask how the lecturer in the ordinary classroom compares with other means of transmitting information. As has already been suggested, if the learner is motivated, a good deal of communication of knowledge apparently can take place with relatively infrequent checks on the progress of the learner and relatively few extrinsic rewards. From Hebb (1949) it could be inferred that relatively fresh ideas would be motivating, but that experiences too far removed from the student's past experience would produce anxiety. This suggests that the organization of materials should be of great importance in learning.

---

[4] D. Beardslee, R. Birney, & W. J. McKeachie, *Summary of Conference on Research in Classroom Processes* (Dept. of Psychology, Univer. of Michigan, 1951) (Unpublished).

Skinner and his students have recently popularized teaching machines which present course materials in well-organized step-by-step fashion with questions (Skinner, 1958). When a student misses a question on some machines, the question is later repeated, but when the correct response is made, the learner proceeds. It seems that the teaching machine should have advantages over the lecturer, for the sequence can be carefully planned to utilize research on the method of successive approximations, on concrete to abstract sequences in problem solving, and on building up generalizations from varying specifics. Lecturers, on the other hand, vary greatly in the degree to which they organize their materials systematically. Probably few lecturers use optimal sequences of presentation.

Moreover, the learner in a lecture is largely passive, while the learner using a teaching machine is continually active. Many studies of different types of learning and concept-formation demonstrate that active learning is more effective than passive learning (e.g., Ebbinghaus, 1913; Wolfle, 1935). Further, the lecturer presents material symbolically. For learning involving perceptual or motor responses, verbal descriptions should be less effective than actual sensorimotor experiences.

Because the lecture provides little feedback, does not always present material in an optimum sequence, allows the student to be passive, and provides little direct experience, lectures may be inferior to other teaching media in achieving certain goals. In fact, if instructors become extinct, they will probably first disappear from the lecture hall.

The chief competitor of the lecturer is not the teaching machine, television, or film, but a much older invention—printing. If rate of transmission of knowledge is important in teaching, a good book is hard to beat. Not only can the reader control his own rate, but the motivated, skilled reader can traverse the printed page much more rapidly than even the fastest lecturer can deliver the material. Over a generation ago, Greene (1928) demonstrated that college students learned as much from reading a passage as from hearing the same material in a lecture.

Although printed materials have been almost as popular as television and have existed for a much longer time, lectures have survived. Even the advent of picture-book textbooks did not dislodge the lecturer. If we stop to think about this, we probably should not be surprised that dozens of research studies have not had much impact upon lecturers' attitudes toward television.

Perhaps the lecturer's arguments are rationalizations, for there is little research to support them. Nevertheless, psychologists may have underestimated important factors in their usual analyses of the learning situation. To maintain good experimental controls, we carefully control rate and sequence of presentation in most of our experiments. The materials used are meaningless to the learner. The results lead us to stress the importance of feedback to the learner.

Lecturing, however, is largely devoted to communicating meaningful materials to somewhat motivated learners. Apparently such learning can take place with relatively infrequent checks on the progress of the learner. In fact, the learner can to some extent obtain feedback by himself. Our experimental controls lead us to overlook the important fact that when knowledge is presented by a teacher he is able to respond to feedback from the learners. This may be an important asset of the instructor. Films and television present material at a relatively fixed rate, but an instructor can go faster or slower when he gets cues of inattention, irritation, or confusion from his class.

The reader too can pace himself, but the inexperienced student may not be able to separate the important from the unimportant. Even though lecturers speak more slowly than books can be read, a good lecturer may be able to give his students all they need from a book in much less time than it would take them to read it.

Textbooks, films, and teaching machines must be organized to fit average students if they are to be economically feasible. The lecturer can not only plan his lecture for his own class but he can respond to feedback from his class as he delivers it. This responsiveness to cues from the class is probably the reason that material can be covered less rapidly in "live" classes than in television classes. Because the instructor responds to feedback, his presentation may appear to be unorganized. Yet one might hypothesize that this very responsiveness makes for greater effectiveness than does a carefully organized, inflexible presentation.

Although there is little relevant research evidence on this matter, we would expect live lecturing to be most effective in situations where there is considerable variation between groups in ability, relevant background, or motivation and where flexible adjustment to the group is therefore important.

Most lecturers avow aims beyond transmission of information. College instructors often say that they provide the integration lacking in the text. Again one would expect that other means of communication could also provide integration. Probably what the instructor really does is to provide his own system of integration. Whether or not this is preferable to the integration provided by a textbook, acceptance of the frame of reference of the instructor does at least make a difference in the grade received by the student. Runkel (1956) measured the structure of instructors' and students' attitudes in beginning college courses in psychology and zoology. He found that *agreement* with the instructor's position *did not* predict students' grades, but students whose attitudes were *colinear* (i.e., who used the same dimensions) with the instructor *did* earn higher grades. In short, students who structured the field as the instructor did tended to earn higher grades. What we do not know yet is whether or not the instructor can communicate his structure to students who do not already have it.

Probably the most careful attempts to measure attitudinal and motivational outcomes have been those comparing live instruction with television instruction in the research programs at Pennsylvania State University and Miami University. In neither case did the live instructor seem to be superior. Still, if the students' tendency to identify with the instructor has anything to do with personal interaction with the instructor, it may be ominous that students did not seek personal conferences with television instructors as much as with "live" instructors.

Methods of Lecturing

Few experiments have compared the effectiveness of classroom lectures with other teaching methods in achieving attitude change of attitudes, we find a substantial and periments to other research dealing with change of attitudes, we find a substantial and growing literature relevant to differing techniques of lecturing.

The research of Hovland and his associates at Yale (1953) indicates that such variables as credibility of the lecturer, order of presentation, presentation of one side of an issue versus presentation of both sides, and emotionality of argument are factors in determining the effect of a lecture. For example, the Hovland group found that a group of college students were more likely to change their opinions (at least temporarily) when they received a persuasive communication from a source which they considered highly credible than when the same communication came from a less credible source. Although we assume that students perceive their instructors as credible sources, faculty rating scales including an item such as "Knowledge of Subject Matter" have revealed that students do discriminate between professors on this dimension. It has been argued, against inclusion of this item, that students are not competent judges of the instructor's knowledge. Regardless of the validity of the student ratings, however, they may indicate the students' credence in the instructor's statements and thus his effectiveness in bringing about attitude changes.

In most of the Yale experiments comparing communication sources with differing credibilities, the actual information gained from the communication was not different for different communicators. In one study, however, a communication from a neutral source resulted in greater informational learning than did the same communication emanating from either an expert or a biased source. Should teachers try to be neutral laymen?

Sometimes even college professors try to reinforce their points by waving the flag or preaching academic hell-fire and brimstone. The Yale studies suggest that this type of teaching is ineffective, for they indicate that the greatest change in reported behavior occurred in those groups to which a minimally fear-arousing lecture was given. This difference persisted at least a year, even though all presentations resulted in equal factual learning.

In presenting controversial points the lecturer often wonders whether he should present the evidence on both sides of the issue or simply present that favoring the position he accepts. Aside from the ethical problems involved, the Yale experiments indicated the greater effectiveness of presenting both sides for (1) an intelligent audience, (2) those initially disagreeing with the lecturer's position, and (3) those who come into contact with the opposing arguments in some setting other than the lecture.

If he wishes to present two sides of an issue, the lecturer is faced with the problem of choosing the order in which to present them. An extensive series of studies has illuminated the effects of differing orders. The most relevant results have been summarized as follows:

> *When contradictory information is presented in a single communication, by a single communicator, there is a pronounced tendency for those items presented first to dominate the impression received* (Hovland, 1957, p. 133).
>
> *The primacy effect found in presenting contradictory information in the same communication was reduced by interpolating other activities between the two blocks of information and by warning the subjects against the fallibility of first impressions* (Hovland, 1957, p. 134).
>
> *Placing communications highly desirable to the recipient first, followed by those less desirable, produces more opinion change than the reverse order* (Hovland, 1957, p. 136).
>
> *When an authoritative communicator plans to mention pro arguments and also nonsalient con arguments the pro-first order is superior to the con-first order* (Hovland, 1957, p. 137).

Research on organization of materials is also relevant to lecturing aimed at cognitive changes. In organizing a lecture the professor frequently is guided by the maxim "Tell them what you're going to tell them. Tell them. Then tell them what you've told them." In a classroom experiment in a course in physics (Lahti, 1956), the instructor first stated a principle and then illustrated and applied it. He compared this with a technique in which he developed principles by demonstration and analysis of application situations and then stated the principle. For students with poor backgrounds the results showed the latter (inductive) method to be superior to the former method on tests of ability to apply the principles. On the other hand, Hovland and Mandell (1952) found that opinions were more likely to change in the direction advocated by the speaker if he drew the conclusion appropriate to his argument rather than leaving it to the students. As we shall again see later in our consideration of laboratory teaching, research on this topic is not clear-cut. Probably it is important for the instructor to point out the conclusion if most students would not arrive at it by themselves. On the other hand, as we have seen, discovery of a conclusion by oneself may have considerable motivational value for a student.

Katona's (1940) classic studies of college student learning supported the importance of organization in learning and retention and also pointed toward the importance of the learner's own organization. Katona found that learning by organization results both in superior retention and in superior application when compared with learning by rote memorization. Such findings make apparent some of the ways in which lecturers are not as effective as they might be.

Size of Lecture Class

Our examination of research on class size led to the conclusion that more meaningful research on this variable must take into account the methods used in teaching classes of differing sizes. Thus, it seems illogical to treat class size separately from method; we will consider class size under the general topics of lecture and discussion.

Among the earliest and most comprehensive programs of research on college teaching were the monumental studies of class size conducted at the University of Minnesota in the 1920's. These studies pointed to the conclusion that large classes are actually superior to small classes. Fifty-nine well-controlled experiments were reported by Hudelson (1928). These experiments involved such widely varying subject matter as psychology, physics, accounting, and law. In 46 of the experiments, results favored the large classes. Although many of the differences were not statistically significant, the majority of significant differences favored large classes. In these experiments small classes averaged 25 to 30 students, but they ranged from 12 to 60 in size while large classes ranged in size from 35 to 150. Extreme differences in size were no more favorable to small groups than were small differences. Although most of the criterion measures were tests of knowledge, some experiments also attempted to measure higher-level intellectual changes, with similar results.

More recent experiments have been less favorable to large classes. Rohrer (1957) found no significant differences. The Macomber and Siegel experiments at Miami University (1956, 1957a, 1957b, 1960) are particularly important because their measures included, in addition to conventional achievement tests, measures of critical thinking and problem solving, scales measuring stereotypy of attitudes, and tests of student attitudes toward instruction. Although the only statistically significant differences favored the smaller classes (particularly for high-ability students), most differences were very small. Significant differences favoring small classes were found on measures of change in misconceptions in psychology, on a case test of problem solving in marketing, and on measures of student attitudes toward all the courses. When retention of knowledge was measured one to two years after completion of the courses, large classes did not prove to be significantly inferior to small classes in any one course. However, in eight of the nine courses compared, differences favored the small class (Siegel, Adams, & Macomber, 1960). As we saw earlier, Nachman and Opochinsky (1958) found a small class to be superior to a large class in performance on surprise quizzes, but the two classes were not significantly different on the final examination for which students prepared.

At Grinnell College, students gave instructors higher ratings in smaller classes (Lovell & Haner, 1955); at Brooklyn (Riley, Ryan, & Lifshitz, 1950) and Purdue (Remmers, 1927), there was no significant difference in ratings of instructors by small and large classes generally, although Remmers (1933) reported that students involved in a controlled experiment at Purdue preferred a small recitation to a large lecture. The weight of the evidence seems to favor small classes if one uses student or faculty satisfaction as a criterion.

Despite the lack of conclusive experimental support for their position, most college professors still believe small classes to be superior to larger ones. For example, the Miami University professors involved in the Macomber-Siegel experiments felt that large classes were about equal to small classes in covering content but inferior in achieving other objectives. In view of the research results, is this simply academic featherbedding or are there good reasons for the professors' distrust of the research results?

Let us briefly return to theory. We have stressed the role of the lecturer as an information communicator. Insofar as information-giving is a one-way process, size of group should be limited only by the audibility of the lecturer's voice. In fact, as Hudelson suggests, a large class may have sufficient motivational value for an instructor to cause him to spend more time in preparation of his lectures and thus to produce better teaching and greater student achievement. But if, as we suggested in our discussion of automation, the effective lecture involves interaction between instructor and students, the large class may be inferior even for lectures, for laboratory experiments suggest that fewer students raise questions or interpose comments in large classes than in small (J. R. Gibb, 1951).

If there is less participation in large classes, the Minnesota research may suggest that we hark back to the criterion problem mentioned earlier. Were the achievement tests used biased against teaching which introduced varying points of view? If our tests place a premium upon exact recall of the materials presented by the teacher or textbook, the student who hears other points of view may be at a disadvantage.

To sum up: large lecture classes are not generally inferior to smaller lecture classes if one uses traditional achievement tests as a criterion. When other objectives are measured, large lectures tend to be inferior. Moreover, both students and faculty members feel that teaching is more effective in small classes. Regardless of the validity of these feelings, any move toward large classes is likely to encounter strong resistance. Probably of more significance than class size per se is its relation to the teaching method used. For example, one would expect class size to be of minimal relevance in television teaching, of slight importance in lecturing, but of considerable significance in discussion teaching. One unplanned consequence of increasing class size may be a restriction upon the teacher's freedom to vary his methods to fit his objectives.

What is the role of the lecturer in higher education? The research results we have cited provide little basis for an answer. They do not contradict—sometimes they support—our earlier notions that the lecture is an effective way of communicating information, particularly in classes where variations in student background, ability, or interest make feedback to the lecturer important. We have also seen that the organization and presentation of lectures may influence their effectiveness in teaching students how to apply knowledge or in influencing attitudes. However, a suspicion, supported by bits of evidence, arises that other methods of teaching may be more effective than lecturing in achieving some higher cognitive and attitudinal objectives.

DISCUSSION METHODS

In our review of research comparing the effectiveness of lecture and discussion methods, we implied that discussion may be ill-suited for communicating information because the rate of communication from instructor to student is slow. However, we should point out that not all information is eagerly received. When information encounters intellectual or emotional resistance, discussion methods may be necessary to bring the source of resistance to light.

Moreover, if students are to achieve application, critical thinking, or some other higher cognitive objective, it seems reasonable to assume that they should have an opportunity to practice application and critical thinking and to receive feedback on the results. Group discussion provides an opportunity to do this. While teaching machines and mock-ups may also be designed to provide prompt and realistic feedback, a group discussion permits presentation of a variety of problems enabling a number of people to gain experience in integrating facts, formulating hypotheses, amassing relevant evidence, and evaluating conclusions. In fact, the prompt feedback provided by the teaching machine may actually be less effective than a method in which students are encouraged to discover solutions for themselves with less step-by-step guidance (Della-Piana, 1957). Since problem solving ordinarily requires information, we might expect discussion to be more effective for groups with more information than for those lacking in background. Some support for this hypothesis is provided by a study of the learning of children in visiting a museum. Melton, Feldman, and Mason (1936) found that lectures were more effective than discussions for children in Grades 5, 6, and 7, but discussions were more effective for eighth-graders.

We noted that lectures usually place the learner in a passive role and that passive learning is less efficient than active. We would expect discussions to promote more active learning, and for once we have some relevant evidence. Bloom and his colleagues (1953) at the University of Chicago used recordings of classes to stimulate students to recall their thoughts during class. As predicted, they found that discussion did stimulate more active thinking than did lecture classes. Krauskopf (1960) substituted written for oral responses to the tape recordings and found that rated relevance of thoughts was positively correlated with achievement, accounting for variance in achievement beyond that accounted for by ability.

The idea that discussion methods should help overcome resistance to learning is also difficult to verify. Essentially, the argument is that some desired learning encounters emotional barriers which prevent it from affecting behavior. For example, a psychology student may learn that distributed practice is effective, but not change his study methods because his anxiety about grades is so great that he doesn't dare try anything different. In such circumstances, experiments on attitude change suggest that the instructor must either bring about changes in underlying attitude and motivation or change the individual's perception of the instrumental relationship between his belief and his motives. Psychotherapists believe that expressing one's attitude in a nonthreatening situation is one step in the process of change. A group discussion may provide such opportunities for expression as well as give opportunities for other group members to point out other instrumental relationships.

In addition, most attitudes influencing learning have some interpersonal antecedents and are stabilized by one's perception of the attitudes of liked persons. Group discussion may facilitate a high degree of liking for the instructor and for other group members. It also permits more accurate assessment of group norms than is likely to occur in other techniques of instruction. Consequently, change may follow.

In fact, while individual instruction would be advantageous for many teaching purposes, group processes can provide a real advantage in bringing about changes in motivation and attitudes. Lewin (1952) showed in his classic experiments on group decision that it is sometimes easier to change a group than an individual.

Whether or not discussions actually are superior in these respects cannot be easily determined, for discussions range from monologues in which occasional questions are interposed to bull sessions in which the instructor is an interested (or bored) observer. Nevertheless, a good deal of research has been attempted to compare the effectiveness of differing discussion techniques.

Student-Centered versus Instructor-Centered Teaching

The theories of client-centered counseling and of Lewinian group dynamics have led to an increased interest in discussion techniques. A wide variety of teaching methods are described as "student-centered," "nondirective," "group-centered," or "democratic" discussion. Nevertheless, they have in common the desire to break away from the traditional instructor-dominated classroom and to encourage greater student participation and responsibility. In Table 1 are listed some of the ways in which the student-centered method has been supposed to differ from the traditional "instructor-centered" method.[8]

With the instructor's role as information-giver reduced, his role as source of feedback virtually eliminated, his opportunity to provide organization and structure curtailed, it is apparent that a heavy burden falls upon the group members to carry out these functions. We expect that these functions could best be assumed by groups which not only have some background in the academic discipline involved but also have had experience in carrying out these functions in "democratic" groups.

TABLE 1

DIMENSIONS UPON WHICH STUDENT-CENTERED AND INSTRUCTOR-CENTERED METHODS MAY DIFFER

Student-Centered	Instructor-Centered
Goals	
Determined by group (Faw, 1949)	Determined by instructor
Emphasis upon affective and attitudinal changes (Faw, 1949)	Emphasis upon intellectual changes
Attempts to develop group cohesiveness (Bovard, 1951b)	No attempt to develop group cohesiveness
Classroom Activities	
Much student participation (Faw, 1949)	Much instructor participation
Student-student interaction (McKeachie, 1951)	Instructor-student interaction
Instructor accepts erroneous or irrelevant student contributions (Faw, 1949)	Instructor corrects, criticizes, or rejects erroneous or irrelevant student contributions
Group decides upon own activities (McKeachie, 1951)	Instructor determines activities
Discussion of students' personal experiences encouraged (Faw, 1949)	Discussion kept on course materials
De-emphasis of tests and grades (Asch, 1951)	Traditional use of tests and grades
Students share responsibility for evaluation (Ashmus & Haigh, 1952)	
Instructor interprets feelings and ideas of class member when it is necessary for class progress (Axelrod, 1955)	Instructor avoids interpretation of feelings
Reaction reports (Asch, 1951)	No reaction reports

From the standpoint of theory, student-centered teaching in its more extreme forms might be expected to have some serious weakness, at least in achieving lower-level cognitive goals, such as knowledge of facts.

Since student-centered teaching attempts to reduce dependence upon the instructor, it would be expected to diminish his influence as a prestige figure, and possibly to reduce his power to bring about attitudinal changes. However, this may be more than compensated for by increased freedom of expression and increased potency of group norms as sources of influence. Participation in discussion gives students an opportunity to gain recognition and praise which should, according to learning theory, strengthen motivation. Thistlethwaite (1960) found that National Merit Scholars checked as one of the outstanding characteristics of teachers who contributed most to their desire to learn, "allowing time for classroom discussion." Other characteristics mentioned include "modifying course content to meet students' needs and interests," "treating students as colleagues," and "taking a personal interest in students." However, in line with our earlier discussion of feedback, another trait mentioned was "providing evaluations reassuring the student of his creative or productive potentialities."

---

[8] A good summary and critical evaluation of studies in this area may be found in R. C. Anderson (1959).

The advocates of student-centered or group-centered teaching also introduced another category of objectives not usually considered in traditional classes—development of skills in group membership and leadership. The group-centered teacher might often argue that even if his method were no more effective than traditional methods in achieving the usual course objectives, it is so important that students learn to work effectively in groups that it may even be worth sacrificing some other objectives in order to promote growth in this area.

Student-centered teachers often stress group cohesiveness. Hence, a possible explanation for the contradictory results on student achievement in the experiments considered below may be found in the studies of group cohesiveness and productivity in industry (e.g., Seashore, 1954). These studies indicate that it is not safe to assume that a cohesive group will be a productive one. Cohesive groups are effective in maintaining group standards, but may set either high or low standards of productivity. Since strongly cohesive groups feel less threatened by management than do less cohesive groups, it may be more difficult to change their standards. Thus in creating "groupy" classes an instructor may sometimes help his students develop strength to set low standards of achievement and maintain them against instructor pressures, or at least to develop group goals different from their normal academic goals.

Another factor may account for the fact that some researchers report student enthusiasm for student-centered teaching while others report hostility. Horwitz (1958) found that aggression toward the teacher increased when the teacher exercised his authority arbitrarily, i.e., refused to abide by the students' decision about teaching methods after telling them that their vote would count. The same method was not resented when the instructor indicated that he would make the final decision. This is important because the limits of student power are often ambiguous in student-centered classes. Also relevant was Horwitz's finding that the more hostility was inhibited, the less learning took place on a type-setting task.

Experiments on student-centered teaching. With this introduction, let us review experimental attempts to demonstrate the effectiveness of student-centered teaching.[*] One of the best known comparisons of student-centered and instructor-centered instruction is that made by Faw (1949). Faw's class of 102 students met two hours a week to listen to lectures and two hours a week in discussion groups of 34. One discussion group was taught by a student-centered method, one by an instructor-centered method, and one group alternated between the two methods.

As compared with the instructor-centered class, the student-centered class was characterized by more student participation, lack of correction by the instructor of inaccurate statements, lack of instructor direction, and more discussion of ideas related to personal experiences.

Faw's major measure of attainment of objectives was in the intellectual area. Scores on the objective course examination based on the textbook showed small but significant differences favoring the student-centered method. In the area of his major interest, emotional growth, Faw's method of evaluation was to ask students in the student-centered and alternating method classes to write anonymous comments about the class. Generally Faw thought that these comments indicated that the students felt that they received greater social and emotional value from the student-centered discussion groups than from an instructor-centered class.

---

[*] Much of the following material on student-centered discussion is based on Birney and McKeachie (1955) and is used with the permission of the American Psychological Association.

A very similar experiment was carried out by Asch (1951). Like Faw, Asch taught all the groups in his experiment. Three sections of about 30–55 students were taught by an instructor-centered method which was half lecture and half discussion. One section of 23 students was taught by a nondirective method quite similar to that of Faw. However, there were certain differences between the two experiments. In Faw's experiment both student-centered and instructor-centered classes spent two hours a week listening to lectures. While Faw did not mention grading, one assumes that grades were determined by the instructor on the basis of the course-wide examination. In Asch's experiment, students in the student-centered class were allowed to determine their own grades.

Asch's results do not completely agree with Faw's. On the final examination in the course, students in the instructor-centered class scored significantly higher than members of the student-centered class on both the essay and objective portions of the test. Note, however, that the student-centered class was specifically told that this examination would in no way affect their grades in the course, and therefore the two groups were probably not equivalent in motivation. Haigh and Schmidt (1956), however, found no significant difference in a similar comparison.

As measured by the Bogardus Social Distance scale, attitude change in the two sections was not significantly different. Yet, in comparison with the instructor-centered class, a greater percentage of members of the student-centered class improved in adjustment as measured by the Minnesota Multiphasic Personality Inventory.

Asch's students, like Faw's, had a different perception of their achievement from that shown by the course examination. Faw's student-centered class did better on the course examination than the instructor-centered class section but thought they would have learned more if they had been in an instructor-centered class. Asch's students, however, rated the student-centered class higher than the instructor-centered class in helping them to learn the subject matter of the course even though they actually scored lower than the instructor-centered class.

Following the model of Lewin, Lippitt, and White's (1939) study of authoritarian, democratic, and laissez-faire group climates, the staff of the University of Michigan's general psychology courses set up an experiment using three styles of teaching: recitation, discussion, and group tutorial (Guetzkow, Kelly, & McKeachie, 1954). As compared to discussion and tutorial methods, the more autocratic recitation method proved not only to produce superior performance on the final examination, but also to produce greater interest in psychology, as measured by the election of advanced courses in psychology. Furthermore, students liked the recitation method better than the other methods. The greater gains in knowledge produced by the recitation method fit in with the general principle that feedback aids learning, for students in the recitation sections had weekly or semiweekly quizzes. McKeachie (1951) suggests that the popularity of this method is related to student anxiety about grades, which is most easily handled when students are in familiar, highly structured situations.

Another factor in these results may be the inexperience of the instructors involved, most of whom had had less than a year of previous teaching experience. It is possible that skillful discussion or tutorial teaching requires greater skill than do more highly structured methods.

Despite the superiority found for the recitation method in producing achievement measured immediately after the course, two results dealing with motivational outcomes favored the other methods. The discussion sections were significantly more favorable than the other groups in their attitude toward psychology; a follow-up of the students three years later revealed that seven men each from the tutorial and discussion groups majored in psychology and none of those in the recitation sections did so.

One of the most comprehensive experiments on student-centered teaching was that of Landsman (1950). He contrasted a student-centered type of teaching with a more direct type of democratic discussion organized around a syllabus. His experimental design involved eight classes in a course sequence of "Human Development," "Adjustment," and "Learning." Three instructors took part in the experiment, and each instructor used both methods. Outcome measures included the Horrocks-Troyer test (an analysis of case histories), a local case-history analysis test, the group Rorschach, the MMPI, autobiographies, and students' reactions. His results showed no significant difference between methods on any of the measures.

Other experiments have also been carried out with negative results. For example, Johnson and Smith (1953) also found no significant difference in achievement test scores between small "democratic" and large lecture classes. An interesting sidelight of their experiment was that one democratic class evaluated their procedure very favorably, while the other democratic class tended to be less satisfied than lecture classes. Bills (1952) also found no difference in achievement between psychology classes taught by lecture-discussion versus student-centered methods, but did find that the students in the student-centered class were significantly more favorable in their attitude toward psychology.

Maloney (1956) found no differences in achievement between two types of discussion groups but did find gains in group cohesiveness, participation, and other indices of effectiveness in groups in which the leader specifically tried to establish acceptance and other characteristics of the student-centered group. In a graduate course in counseling, Slomowitz (1955) found no significant differences between nondirective and problem-oriented discussions in achievement or application to a case-study test. The nondirectively taught student did, however, change in self concept more (at the .10 level of confidence) than did students in the other group. Like Slomowitz, Deignan (1955) was concerned not only with achievement but also with emotional changes. Although there were no other significant differences between the groups on follow-up tests a semester later, the student-centered group evaluated the course more highly. Rasmussen (1956) also reported no difference in achievement but higher morale among groups of in-service teachers electing student-centered extension courses as compared with students taught by instructor-centered methods. Krumboltz and Farquhar (1957) compared student-centered, instructor-centered, and eclectic teaching methods in a how-to-study course. There were no significant differences between methods on an achievement test or on self-ratings of study habits. Burke's results (1956) are similar.

In an attempt to teach critical thinking, Lyle (1958) compared a problem-oriented approach with conventional lecture-discussion-text procedures. Apparently he found that the conventional group was superior to the problem-oriented group in achievement. Gains in critical thinking were not greater in the problem-centered classes. When students were asked to write a question for the final examination, the conventional group wrote factual questions and the problem-centered group wrote "thought" questions.

In an experiment in teaching written and spoken English, Jenkins (1952) found that a teaching role in which the instructor acted as resource person in a democratic class was not significantly superior to traditional methods on three measures of communication skills.

The Johnson and Smith study (1953) is one of the few to support our earlier suggestion that the success of student-centered teaching depends upon the previous group experience of the students. They suggest that the critical factor in the success of one democratic class was the enthusiastic participation of one student who happened to be a member of a student cooperative.

Wispé (1951) carried out an interesting variation of the student-centered versus instructor-centered experiment. Instead of attempting to control the instructor personality variable by forcing instructors to teach both instructor-centered and student-centered classes, Wispé selected instructors who were rated as naturally permissive or directive. He then compared their sections of the Harvard course in "Social Relations." He found no difference in final examination scores between students taught by the different methods. Students preferred the directive method, and the poorer students gained more in directive classes.

In a methodological aside, it is worth noting that analysis of covariance and multiple correlation were used in Wispé's study. As indicated in our introduction, one reason why statistically significant differences are so rare in this research may be that the researchers commonly use "weak" statistics in which the error estimates are inflated by failure to take account of known sources of variance such as individual differences in previous knowledge or intelligence.

As a counterpoint to Wispé's study with teachers using their preferred method, we have the Springfield College experiment (Ashmus & Haigh, 1952; Haigh & Schmidt, 1956) in which the students were given their choice of group-centered or instructor-centered teaching. Students choosing either type did not differ significantly in intelligence. The results showed no difference in the achievement of the two groups on a test not counted toward the course grade. Those in the nondirective classes were more highly satisfied with the course.

Moore and Popham (1959) reported that three student-centered interviews with students produced greater gains on the college Inventory of Academic Adjustment than did three content-centered interviews conducted outside of class in an educational psychology course. Similarly, Zeleny (1940) found not only greater gain in knowledge but also greater self-rated change in personality in a group-discussion class than in a traditional recitation-discussion class.

While scores on objective final examinations seem to be little affected by teaching method, there are, in addition to the changes in adjustment reported by Asch, Faw, Zeleny, and Moore and Popham, other indications that student behavior outside the usual testing situation may be influenced in the direction of educational goals by student-centered teaching. The classes compared by Bovard (1951a, 1951b) and McKeachie (1951) differed in the degree to which interaction between students was encouraged and in the degree to which the class made decisions about assignments, examinations, and other matters of classroom procedure. Like other experimenters, Bovard and McKeachie found that the groups did not differ in achievement measured by the final examination. However, two clinical psychologists evaluated recordings in the class discussions which followed the showing of the film, "The Feeling of Rejection." Both clinicians reported that the "group-centered" class showed much more insight and understanding of the problems of the girl in the film.

Patton (1955) felt that an important variable in group-centered classes was the students' acceptance of responsibility for learning. In his experiment he compared traditional classes with two classes in which there were no examinations, no lectures, and no assigned readings. Students in the experimental classes decided what reading they would do, what class procedure would be used, what they would hand in, and how they would be graded, so that they had even more power than had previous experimental groups. At the end of the course, these classes, as compared to the control group, felt that the course (1) was more valuable, (2) showed greater interest in psychology, and (3) tended to give more dynamic, motivational analyses of a problem of behavior.

But giving students power cannot work if they will not accept responsibility; so Patton also obtained individual measures of acceptance of responsibility within the experimental classes. As hypothesized, he found that the degree to which the student accepted responsibility was positively correlated with gain in psychological knowledge, gain in ability to apply psychology, rating of the value of the course, and interest in psychology. The effect of giving students additional responsibility seemed to depend upon the student's readiness to accept that responsibility.

Although the Bovard, McKeachie, and Patton experiments suggested that student-centered classes promoted transfer of learning, D. E. P. Smith (1954), comparing three methods varying in degree of directiveness, found no differences in their effects upon students' abilities to make "applicational transfer" of their learning.

The most impressive findings on the results of small-group discussion come from the research on the "Pyramid Plan" at Pennsylvania State University (Carpenter, 1959; Davage, 1958, 1959). The basic plan may be represented by a description of their experiments in psychology. Each "Pyramid Group" of psychology majors consisted of six freshmen, six sophomores, two juniors, who were assistant leaders, and a senior, who was group leader. The Pyramid groups met weekly for two-hour periods to discuss personal-professional goals, requirements for entering their selected professions, skills needed for academic success, the significance of their courses for their goals, issues in higher education, and the central concepts of psychology. The group leaders worked with a faculty supervisor in defining objectives, discussing small-group theory and techniques, and defining issues to be considered by the Pyramid groups. One control group consisted of students who simply took pretest measures; another control group received comparable special attention by being given a special program of lectures, films, and demonstrations equal to the time spent in discussion by the Pyramid groups. The results on such measures as attitude toward psychology, knowledge of the field of psychology, scientific thinking, use of the library for scholarly reading, intellectual orientation, and resourcefulness in problem solving were significantly favorable to the Pyramid Plan. Moreover, a follow-up study showed that more of the Pyramid students continued as majors in psychology. Such an array of positive results, little short of fantastic, testifies not only to the effectiveness of the Pyramid program but also to the resourcefulness of the Pennsylvania State research staff.

L. M. Gibb and Gibb (1952) reported that students who were taught by their "participative-action" method were significantly superior in role flexibility and self-insight to students taught by traditional lecture-discussion methods. In the participative-action method class, activities centered around "subgrouping methods designed to increase effective group participation."

> The instructor, who played a constantly diminishing role in the decisions and activities of the groups, gave training in role playing, group goal setting, problem centering, distributive leadership, evaluation of individual performance by intra-group ratings, process observing, and group selection, evaluation, and revision of class activities (L. M. Gibb & Gibb, 1952, p. 247).

Gibb and Gibb also supported the assumption that group-centered teaching can facilitate development of group-membership skills. They found that in nonclassroom groups the participative-action students were rated higher than other students in leadership, likeableness, and group-membership skills. DiVesta's results tend to support this (1954), and R. P. Anderson and Kell (1954) reported that members of student-centered groups were characterized by positive attitudes toward themselves as participants in the group.

Although McKeachie (1954) reported significant changes in attitudes of students toward Negroes and toward the treatment of criminals, differences between leader-centered and group-centered classes were not significant. As predicted, however, group decision did produce more accurate perception of the group norm and more conformity to the norm than lecture or discussion without decision. While no direct attempt was made to change the group norm, the experiment suggests that the instructor who wishes to change attitudes might find the group-decision technique useful. Wieder (1954) found that a nondirectively taught psychology class tended to reduce prejudice more than did conventional classes.

One final bit of support for less directive teaching is Thistlethwaite's (1959) finding of a significant negative correlation between a college's productivity of Ph.D.'s in natural science and the directiveness of the teaching methods used.

In the Guetzkow, Kelly, and McKeachie (1954) experiment reported earlier, none of the males in the very popular highly structured recitation sections majored in psychology. At least seven males in the sections taught by each of the other two methods completed a psychology major. Since the three methods had been equated in terms of the number of students intending to major, this is a surprising finding. Such results are shocking to conventional theory. Most psychologists have thought that prompt feedback and well-structured sequences of presentation were conducive to learning. Yet the Thistlethwaite studies indicate that the top colleges in production of scholars are those where tests are infrequent and where students don't know what to expect next. If this is so, maybe we need to throw some random elements into our lectures and teaching machines. In any case, these results suggest that techniques most suitable for teaching knowledge may not be those most effective for developing motivation and high-level achievement.

The results of the spate of research on student-centered teaching methods are not conclusive, but tend to support the theory with which our discussion began. We had suggested that student-centered teaching might be ineffective in achieving lower-order cognitive objectives. Experiments reporting losses and gains in this area seem to be balanced. Students apparently can get information from textbooks as well as from the instructor.

We had also predicted that any superiority of student-centered discussion methods would be revealed in higher-level outcomes. In 11 studies, significant differences in ability to apply concepts, in attitudes, in motivation, or in group membership skills have been found between discussion techniques emphasizing freer student participation compared with discussion with greater instructor dominance. In 10 of these the differences favored the more student-centered method. The eleventh (Guetzkow, Kelly, & McKeachie, 1954) had mixed results.

In short, the choice of instructor-dominated versus student-centered discussion techniques appears to depend upon one's goals. The more highly one values outcomes going beyond acquisition of knowledge, the more likely that student-centered methods will be preferred.

Variations in Discussion Teaching

Buzz sessions. One popular technique for achieving student participation in groups is the buzz session (McKeachie, 1960). In using this procedure, classes are split into small subgroups for a brief discussion of a problem. Although many teachers feel that this technique is valuable as a change of pace or as a way of getting active thinking from students, little research has tested its effectiveness. Vinacke (1957) found that, in comparison with their performance on a pretest, students in two- and three-man groups wrote more new ideas after a five-minute discussion than did students working alone. It is possible, however, that similar changes could have been produced by a general discussion or a lecture.

Leadership. Laboratory and field studies of group processes may shed some light on factors which condition the effectiveness of groups and which thus may help account for the lack of effectiveness of many discussion classes.

For example, one problem faced by the discussion leader is the student who dominates discussion or the clique who carry the ball. In some discussion classes, the instructor's fear of exerting too much control over the class may result in failure to give minority points of view an adequate hearing. Research suggests that effectiveness of group problem solving depends upon the leader's ability to obtain a hearing for minority opinion (Maier & Solem, 1952).

Some student-centered teachers assume that all decisions should be made by the group. Hence, they feel, the instructor should not determine objectives, make assignments, or give tests, but should ask the group to make these decisions. If the group does this, can the time they lose from academic work in making decisions be compensated for by the increased motivation of the students? Democratic methods permit formation of group norms and more acute perception of group norms, but as in industry these norms may not necessarily be conducive to high productivity or learning. The general question of the areas in which the instructor should make decisions is one which different instructors have answered in different ways and one well worth further discussion and research. One hunch based on research on business conferences is that the instructor should make most procedural decisions, leaving the class time for problems related to the content of the course (Heyns, 1952).

Even in discussion of course content, however, it appears that some instructor direction may be useful if the goals are the learning of relationships and the ability to apply this learning. In comparing groups given more versus less instructor direction in discovering the basis of solutions of verbal problems, Craig (1956) found that the directed group not only learned faster but retained their learning better than the group given less help. This result is supported by Corman's research on guidance in problem solving (1957).

Studies of business conferences have also shown that one of the commonest causes of dissatisfaction is the member's failure to understand the purpose of the conference. It is little wonder that a student with a high need to achieve success may be frustrated and often aggressive in a democratic class in which he is confused about the purposes of the teacher's procedures and, at the same time, subject to the stress involved in getting a good grade. Bloom's studies (1953) of student thought processes show that on the average 30 per cent of the student's time in discussion classes is spent in thinking about himself and other people as compared with 18 per cent of the time in lectures. With members of the group thus concerned about their own needs, it is no wonder that discussion groups are not always productive.

Grading. Another important problem in conducting student-centered (or other) classes is that of grades. Not only does the instructor control the pleasantness or unpleasantness of a good many student hours, but because of his power to assign grades he can block or facilitate the achievement of many important goals. The importance of this aspect of the teacher's role is indicated by studies of supervision in industry. In one such study it was discovered that workers were most likely to ask a supervisor for help if the supervisor was not responsible for evaluating his subordinates (Ross, 1956). This suggests that as long as students are anxious about the grades the instructor will assign, they are likely to avoid exposing their own ignorance.

The student's anxiety about grades is likely to be increased if his instructor's procedures make him uncertain about what he must do to attain a good grade. Although we have already seen that a high degree of structure may not lead to better education, it is nevertheless worth noting that for many students democratic methods seem unorganized and ambiguous. In an ordinary course the student knows that he can pass by reading assignments and studying lecture notes, but in a student-centered class the instructor does not lecture, does not make assignments, and leaves the student ignorant of what the instructor is trying to do.

Some instructors have thought that the grade problem might be solved by using a cooperative system of grading. Deutsch (1949) found no differences in learning between students in groups graded cooperatively and those graded competitively, although the cooperative groups worked together more smoothly. Following up Deutsch's work, Haines (1959) also found no significant achievement advantages for students working cooperatively as compared with those working competitively for grades, but he did find marked differences in group morale. Haines's work suggests that cooperative grading in the discussion group can be successfully combined with individual grading on achievement tests.

In comparing a "teamwork" class using group incentives with a lecture class, H. C. Smith (1955) did not find differences in satisfaction comparable to those found by Haines and by Deutsch in their experiments.

Complicating the problem of grading is the probability that low grades produce different effects upon different students. Waterhouse and Child (1953) found that frustration produced deterioration in performance for subjects showing high "interference tendencies" but produced improved performance for those with low "interference tendencies." "Interference tendencies" were assessed by a questionnaire in which the subjects were asked to check on a six-point scale the degree to which 90 responses to frustration were characteristic of them.

Considering the importance of grading for both students and instructors, it is regrettable that there is so little empirical research on it. How do students learn to evaluate themselves? How do they learn to set goals for themselves? Do differing grading procedures facilitate or block such learning? Can more educational substitutes for grades be devised? To these questions we have no answers at the present time.

Size of Discussion Group

One of the earliest experimental studies on college teaching was that of Edmonson and Mulder (1924) on class size. This study was conducted in an education course in which there were two sections—one of 45 students and the other of 109. Both sections were taught by the same instructor in order to control the possible effect of instructor differences, and both sections used the same text and took the same tests. The discussion method was used in each section. Forty-three students in each class were paired on the basis of intelligence and past experience. This pioneer study led to the conclusion that size of class is not a significant variable in effecting student learning as measured by usual course achievement tests, although students preferred the small class (if 45 can be considered a small discussion group).

As in the size experiments reported in our discussion of the lecture method, there seemed to be little theoretical reason for the choice of class sizes in this experiment. There is, in fact, some doubt whether either size is optimal for discussions. However, similar results were reported by Hudelson (1928) when using classes of 20 and 113 in an education course, and by Brown (1932) in psychology classes. In fact, using special team procedures, Brown produced slightly better achievement in groups of 60 than were obtained from discussion classes of 25.

Support for small classes, however, comes from the studies on teaching French conducted by Cheydleur (1945) at the University of Wisconsin between 1919 and 1943. With hundreds of classes ranging in enrollment from 9 to 33, Cheydleur found the smaller classes to be consistently superior on departmental examinations, with reliabilities of .72 to .95. These departmental examinations correlated about .8 with standardized tests of achievement in French. The only thing unclear in Cheydleur's report is whether or not some selective factor could have been operating.

Mueller (1924) in a pioneer study also found the smaller class to be more effective in an experiment comparing elementary psychology classes of 20 and 40 students. A study by Schellenberg (1959) in a Western civilization course suggested that even the smallest groups in these studies may be above optimal sizes. Working with discussion groups of 4, 6, 8, and 10 students, he found higher satisfaction and higher instructor grading in the smaller groups. While Schellenberg recognized that grades are an unsatisfactory criterion since the instructor's judgment may shift from section to section, he referred to laboratory studies of group problem solving which point to optima in the range of 4- to 6-person groups.

From the standpoint of theory one might expect increasing size to have two effects. One of these would be an increase in the resources of the group in knowledge, different approaches to the problem, and ability to provide feedback. The second consequence of size, however, is likely to be a decreasing ability to exploit the total resources of the group because of the difficulty in obtaining contributions from everyone. Further, with increasing size, group members are likely to feel restraints against participation (J. R. Gibb, 1951). The consequence of increasing feelings of threat in larger groups is that group participation is increasingly dominated by a few people. In Princeton classes of 4 to 12 students, Stephan and Mishler (1952) reported that increasing group size was related to increasing instructor dominance. Thus group size becomes a much more relevant variable in classes taught by discussion than in those taught by lecture.

Homogeneous versus Heterogeneous Grouping

One common criticism of discussion classes is that class time is wasted either by discussion of problems raised by the able students which are beyond the ken of the other students or by problems raised by poor students which the other students feel are too elementary. One answer to such criticism is to use homogeneous groupings so that each student is discussing problems with students of his own ability.

Recently, concern about America's resources of high-level talent has resulted in the proliferation of honors programs featuring homogeneous classes for students with high academic aptitude and achievement. The logic of such programs is evident, and research evidence supports their educational value. Nevertheless, an earlier college experiment on ability grouping showed no significant advantages for homogeneous grouping by intelligence and even a trend toward the opposite result in psychology classes (Longstaff, 1932).

Briggs (1947), on the other hand, found that special intensive seminar classes meeting less often than conventional classes produced greater achievement for superior students than did the conventional class for a control group matched in ability. Unfortunately, in this study the seminar students volunteered and were selected by interview so that they probably had higher motivation than their controls.

The two earliest publications in this area (Burtt, Chassel, & Hatch, 1923; Ullrich, 1926) reported results which seem very reasonable. They both concluded that homogeneous groups were not superior to heterogeneous groups when given standard material but did superior work when the bright students were pushed. Similarly, Tharp (1930) found homogeneous grouping to be superior in foreign language classes and Taylor (1931) found it to be superior in analytic geometry. All in all, it seems safe to conclude that homogeneous grouping by ability is profitable, if teaching makes use of the known characteristics of the groups.

Homogeneous grouping by personality proved to be ineffective in Hoffman's (1959) experiment in group problem solving. Comparing groups of four students who were similar in personality profiles on the Guilford-Zimmerman Temperament Survey with groups made up of dissimilar students, he found that the *heterogeneous* groups produced *superior* solutions. Hoffman accounts for this difference by suggesting that heterogeneous groups are more likely to propose a variety of alternatives permitting inventive solutions.

On the other hand, in a study by Stern and Cope (Stern, 1960, pp. 315–316), groups of "authoritarian," "antiauthoritarian," and "rational" students in a citizenship course were segregated into homogeneous groups in which the instructor was unaware of which group he was teaching. Authoritarian students in the experimental group achieved more than comparable authoritarians did in conventional classes.

It is apparent that we need further analysis to determine what kinds of homogeneities or heterogeneities contribute to what objectives. If we omit from consideration the general adjustment problems of segregated groups, the idea that one should be able to do a better job of teaching a group with known homogeneous characteristics than a heterogeneous group seems so reasonable that it is surprising that the results of research on this idea are not uniform. It may be that the potential advantages of carefully planned grouping have not been realized simply because we have not yet learned optimal teaching procedures for such groups.

From a theoretical point of view the importance of group size and heterogeneity probably depends upon the purpose of the discussion. If, for example, one is interested in using group members as sources of information or differing points of view, the larger and more heterogeneous the group, the greater its resources. On the other hand, the degree to which a group uses its resources depends upon communication; fewer group members can participate in a large group and members are less likely to volunteer their contributions in large groups.

A final problem is that of group pressures toward consensus and conformity. As we saw in our consideration of discussion techniques, one barrier to effective group problem solving is the tendency of a group to accept the majority view without sufficient consideration of minority views. Since group members may be less likely to express divergent opinions in large groups than in small groups, we might venture the paradoxical hypothesis that the larger the group, the more effect a few outspoken members are likely to have in determining the success of discussion.

LABORATORY TEACHING

The laboratory method is now so widely accepted in scientific education that it may seem heretical to ask whether laboratory experience is an effective way to achieve educational objectives.

Laboratory teaching assumes that first-hand experience in observation and manipulation of the materials of science is superior to other methods of developing understanding and appreciation. Laboratory training is also frequently used to develop skills necessary for more advanced study or research.

From the standpoint of theory, the activity of the student, the sensorimotor nature of the experience, and the individualization of laboratory instruction should contribute positively to learning. Information cannot usually be obtained, however, by direct experience as rapidly as it can from abstractions presented orally or in print. Films or demonstrations may also short-cut some of the trial and error of the laboratory. Thus, one would not expect laboratory teaching to have an advantage over other teaching methods in amount of information learned. Rather we might expect the differences to be revealed in retention, in ability to apply learning, or in actual skill in observation or manipulation of materials. Unfortunately, little research has attempted to tease out these special types of outcomes. If these outcomes are unmeasured, a finding of no difference in effectiveness between laboratory and other methods of instruction is almost meaningless since there is little reason to expect laboratory teaching to be effective in simple communication of information.

In a course on methods of engineering,

White (1945) found that students taught by a group-laboratory method achieved more than those taught by a lecture-demonstration method. However, Balcziak (1953), comparing (1) demonstration, (2) individual laboratory, and (3) combined demonstration and laboratory in a college physical science course, found no significant differences between them as measured by tests of information, scientific attitude, or laboratory performance.

Kruglak (1952) compared the achievement of students in elementary college physics laboratories taught by two methods—the demonstration and the individual method. The individual method proved superior not only in student learning of techniques but also in solving simple laboratory problems. Another experiment on methods of laboratory instruction found that a problem-solving method was superior to traditional laboratory-manual methods in teaching students to apply principles of physics in interpreting phenomena (Bainter, 1955). Lahti (1956) also found a problem-solving method to be superior to more conventional procedures in developing ability to design an experiment. Because many laboratory teachers have been interested in teaching problem-solving methods, this may also be an appropriate place to note Burkhardt's finding (195[illegible]) that students who are taught calculus with an emphasis on the understanding of concepts learn concepts better than students taught with conventional emphasis upon solving problems. On the face of it this might appear to be in opposition to the results of Kruglak, Bainter, and Lahti. Actually all of these studies point to the importance of developing understanding rather than teaching students to solve problems by going through a routine series of steps. Whether or not laboratory is superior to lecture-demonstration in developing understanding and problem-solving skills probably depends upon the extent to which understanding of concepts and general problem-solving procedures are emphasized by the instructor in the laboratory situation.

PROJECT METHODS AND INDEPENDENT STUDY

The recent interest in independent study as a means of utilizing faculty time more efficiently has brought to the fore a teaching method which has been used in some form for many years. If one goal of education is to help the student develop the ability to continue learning after his formal education is completed, it seems reasonable that he should have supervised experience in learning independently—experience in which the instructor helps the student learn how to formulate problems, find answers, and evaluate his progress himself. One might expect the values of independent study to be greatest for students with high ability and a good deal of background in the area to be covered since such students should be less likely to be overwhelmed by difficulties.

Independent study programs frequently involve the execution of projects in which a student or group of students undertakes to gather and integrate data relative to some problem.

The results of research on the effectiveness of the project method are not particularly encouraging. One of the first "independent study" experiments was that of Seashore (1928). His course consisted primarily of guided individual study with written reports on eight projects, each of which took about a month to complete. Final examination scores, however, were no different for these students than for students taught by the usual lecture-discussion method (Scheidemann, 1929). In a college botany course, Novak (1958) found that students in conventional classes learned more facts than did those taught by the project method. The project method was particularly ineffective for students in the middle third of the group in intelligence. Similarly, Goldstein (1956) reported that students taught pharmacology by a project method did not learn more than those taught in a standard laboratory.

Unfortunately, measures of achievement such as those used in the studies just noted are probably not sufficient measures of the objectives of project instruction. Presumably the real superiority of the project method should be revealed in measures of motivation and resourcefulness. Novak's experiment was laudable in its inclusion of a measure of scientific attitude, but neither conventional nor project classes made significant gains from the beginning to the end of the semester. Similarly, in a class in mental hygiene, Timmel (1954) found no difference in the effectiveness of the lecture and project methods in changing adjustment. One morsel of support comes from Thistlethwaite's (1960) finding that National Merit Scholars checked requirement of a term paper or laboratory project as one characteristic of their most stimulating course.

With the support of the Fund for the Advancement of Education, a number of colleges have recently experimented with more elaborate programs of independent study. As with other comparisons of teaching methods, few differences have been found between achievement of students working independently and those taught in conventional classes. Moreover, the expected gains in independence have also often failed to materialize. Students taught by independent study do not always seem to develop greater ability or motivation for learning independently.

One of the most comprehensive research programs on independent study was that carried out by Antioch College (Churchill, 1957; Churchill & Baskin, 1958). The Antioch experiment involved courses with varying periods of independent study in humanities, social science, and science. A serious attempt was made not only to measure cognitive and affective achievement but also to evaluate the effect of independent study upon "learning resourcefulness." In addition, the Antioch staff, recognizing that not all students are ready to work independently, planned programs of training for independent work.

The results of the experiments, however, do not point clearly to any conclusion. For example, in one experiment, independent small groups learned more subject matter in physics than students working independently as individuals. But in art, students working individually learned more than those in independent small groups. As in most experiments on teaching methods, the predominant results were of "no significant difference." An exception to this may be found in various indices of student satisfaction in which several significant differences favor lecture-discussion over independent study and especially over independent small groups.

In the Antioch College study, however, one method of saving faculty time was found effective as measured by both student learning and student attitudes. This method, used in teaching French, increased the students' time in class from six to eleven hours a week, but eight of the instructional hours were in charge of a student assistant who used visual materials prepared by the instructor (Baskin, 1960).

In another well-controlled experiment, carried out at Oberlin College (McCollough & Van Atta, 1958), students in introductory science, psychology, and mathematics were required to work independently of the instructor in small groups. This independent work occupied one-third of the college year following several weeks of preliminary training. As in the Antioch experiment, no significant differences in learning appeared either as measured by the usual achievement tests or by a test of learning resourcefulness. Generally, the Oberlin students seem not to have been unhappy about the independent study experience although they indicated that they would have preferred several two-week periods of independent study to the single longer period.

The most favorable results on independent study were obtained at the University of Colorado by Gruber and Weitman (1960). In a course in freshman English in which the group met about 90 per cent of the regularly scheduled hours and had little formal training in grammar, the self-directed study group was significantly superior to control groups on the test of grammar. In a course in physical optics, groups of students who attended class independently of the instructor but were free to consult him learned fewer facts and simple applications but were superior to students in conventional classes in difficult applications and learning new material. Moreover, the latter difference was maintained in a retest three months later while the difference in factual knowledge disappeared (Weitman & Gruber, 1960). In a class in educational psychology an experimental class meeting once a week with the instructor and twice a week in groups of five or six students without the instructor was equal to a conventional class hearing three lectures a week in mastery of content, but tended to be superior on measures of curiosity. In another experiment, students in self-directed study paid more constant attention to a lecture than did students in conventional classes.

The experiment by McKeachie, Lin, Forrin, and Teevan (1960) also involved a fairly high degree of student-instructor contact. Students normally met with the instructor in small groups weekly or biweekly, but they were free to consult the instructor whenever they wished. The results of the experiment suggest that the "tutorial" students did not learn as much from the textbook as students taught in conventional lecture and discussion section classes, but they did develop stronger motivation both for course work and for continued learning after the course. This was indicated not only by responses to a questionnaire administered at the end of the course but also by the number of advanced psychology courses later elected.

Independent study experiments have varied greatly in the amount of assistance given students and in patterning instructional versus independent periods. For example, merely excusing students from attending class is one method of stimulating independent study. The results of such a procedure are not uniform but suggest that classroom experience is not essential for learning. However, the kinds of learning that take place out of class may be different from those that take place in class.

Jensen (1951) compared four groups, including one in which students were completely excused from class attendance. The results showed no difference in gains among the four groups, but students who had worked independently were more willing than others to volunteer for further independent study. Wakely, Marr, Plath, and Wilkins (1960) compared performance in a traditional four-hours-a-week lecture class with that in a class which met only once a week to clear up questions on the textbook. In this experiment the traditional classes proved to be superior. Similarly, Paul (1932) found 55-minute class periods to be superior to 30-minute periods as measured by student achievement.

The results of studies in a child development course by Parsons (1957) and Parsons, Ketcham, and Beach (1958) were more favorable to independent study. In the latter experiment four teaching methods were compared—a lecture, instructor-led discussions, autonomous groups which did not come to class, and individual independent study in which each student was sent home with the syllabus and returned for the final examination. All groups were chosen randomly. In both experiments students working independently made the best scores on the final examination, which measured retention of factual material in the textbook. The instructor-led discussion groups were the lowest in performance on the final examination. There were no significant differences between groups on a measure of attitudes toward working with children. The authors explain their results on the grounds that the independent group was not distracted by the interesting examples, possible applications, or points of view opposing those in the text, all of which were presented in the other groups.

However, in the latter experiment one group of students was made up of teachers commuting to campus for a Saturday class. The results for these students were quite different from those for resident students.

In this case students in independent study performed significantly worse than other groups on the examination, perhaps because they were less committed to regular study and may also have experienced more frustration in not having class.

Although the Parsons, Ketcham, and Beach results were favorable to independent study, they are not satisfying to the advocate of this method. These results lead to the conclusion that if a student knows that he is going to be tested on the factual content of a particular book, it is more advantageous for him to read that book than participate in other educational activities. In fact, one might suggest that even better results could be obtained if the desired facts could be identified by giving the students test questions in advance. But knowledge of specific facts is not usually the major objective of an independent study program. What we are hoping for is greater integration, increased purposefulness, and more intense motivation for further study. That independent study can achieve these ends is indicated by the Colorado and Michigan experiments. But the paucity of positive results suggests that we need more research on methods of selecting and training students for independent study, arranging the independent study experience, and measuring outcomes. Note that the Colorado results came in courses in which a good deal of contact with the instructor was retained.

AUTOMATED TECHNIQUES

The impending shortage of college teachers has sparked several hotly contested skirmishes about the virtues or vices of various techniques of teaching with devices substituting for a portion of the usual face-to-face interaction between instructors and students. Since some college faculty members anxious about technological unemployment resist innovations, research has often been used as a technique of infiltration rather than as a method of developing and testing educational theory.

Along with many inconsequential studies, a few carefully executed programs of research have emerged. Representative of these are those described in the following section.

Television[7]

Before reviewing the research on teaching by television, let us consider two hypotheses that may help in analyzing the research results.

Television is not a method of instruction in the sense that discussion and lecture are methods of instruction. Rather it is a means of giving the student a clear view of the instructional situation. Therefore, we would expect that (1) the relative effectiveness of teaching via television will vary depending on the importance of being able to see clearly. For example, we would expect television to be effective when it is important for students to see demonstrations, visiting lecturers, or films, but to have little advantage when communication is primarily verbal.

Television reduces the opportunity for students to communicate with teachers and for teachers to interact with students. We would thus expect that (2) the effectiveness of television will vary inversely with the importance of two-way communication not only for feedback to the student but particularly for feedback to the teacher.

In 1954, Pennsylvania State University (Carpenter & Greenhill, 1955, 1958) received a grant from the Fund for the Advancement of Education to study the effectiveness of conventional courses taught for a full semester over closed-circuit television as compared with the same instruction given in the usual manner. With these funds, a program of research was initiated on the courses entitled General Chemistry, General Psychology, and Psychology of Marriage. In 1957 the research program was expanded to (1) extend the project to additional courses; (2) study instructional variables, and (3) work on methods of improving instruction on television.

---

[7] Some portions of this section were previously published in W. J. McKeachie, TV for college instruction, *Improving Coll. & Univer. Teaching*, 1958, 6, 84–89, and are used by permission of the publisher.

The results of this research may be used either to extol or damn television. Essentially they indicate that there is little loss in student learning in courses taught by television classes were lower than those of stuventionally. For example, the first experiment dealt with the lecture portion of courses in general chemistry and general psychology. In the chemistry course the differences between methods on objective measures of information were not significant. In the psychology course the conventional class did prove to be slightly superior in knowledge to the television class.

Students learned the information needed to pass examinations, and most did not object strongly to the televised classes although they preferred live instruction. Students in psychology were asked (1) how much they liked psychology, and (2) how much it contributed to their education as compared with other courses they were taking. On both counts, ratings of the students in the television classes were lower than those of students who were in the same room as the instructor. The psychology students were also asked if they would like to take another course in psychology. About the same percentages of students signed up from lectures, television originating rooms, and television receiving rooms, but when asked if they would prefer taking it in a large class or by television, a plurality preferred television. While students at other colleges do not rebel at television either, research findings quite consistently report less favorable attitudes toward courses taught by television as compared with conventional classes (e.g., Lepore & Wilson, 1958; Macomber & Siegel, 1960).

The effectiveness of television is particularly astonishing to one who has seen the apparent inattention in many television classes. If students who are obviously inattentive learn as much as those who appear to be attentive, one might suspect that many students have developed the ability to appear attentive in conventional classes even though their minds are wandering. In the latter case, television students may simply be exposing the reactions to lectures which polite students ordinarily hide from their instructor. The New York University research staff (Klapper, 1958) used student observers to classify television students into high-, middle-, and low-attention groups. In none of their courses was attention level related to achievement, although their evidence suggested that students who had more previous knowledge of the subject paid less attention than those who had less.

Factors unimportant in using television. The heading of this section would normally be "Factors Conditioning the Effectiveness of Educational Television," but the results of the research are more clearly indicated by the title chosen. For example, recognizing that instructor-student interaction is sometimes important in learning, the Pennsylvania State group installed "two-way" microphone communication in the receiving rooms so that students in the receiving rooms could ask questions. This technique has been used even more extensively at the State University of Iowa (Stuit, et al., 1956) and at Case Institute of Technology (Martin, 1957a, 1957b). The Pennsylvania State group found that this method of instruction was not superior to simple one-way communication, although students prefer two-way communication (Carpenter & Greenhill, 1958). Similar results were found in the research on television instruction in the Army and at Case (Fritz, 1952; Martin, Adams, & Baron, 1958).

Another attempt to combine the value of interaction with that of television was an experiment in presenting a 35-minute television lesson followed by a 15-minute discussion *period in each of the receiving rooms. Other* students in the same course observed by means of television the 15-minute discussion conducted by the instructor with the eight students in the origination room. Still other students were allowed to leave or to study their notes. As in the other attempt to provide interaction, results showed no significant differences in test performance between students taught by each of these three methods. A poll of students indicated that they preferred two hours of lecture followed by a full period of discussion to a short discussion each period.

Size of the viewing group is not an important variable in television instruction, nor do proctors in the viewing rooms contribute to student learning. Adapting a course for television by adding supplementary visual aids also proved to be no more effective than televised lecture-blackboard presentations (Carpenter & Greenhill, 1958). In fact, both at Pennsylvania State and at New York University (Adams, Carpenter, & Smith, 1958), the "visual" productions tended to be *less* effective than "bare bones" television. This result should probably not startle us after having read the Parsons, Ketcham, and Beach results (1958) in research on independent study. Just as discussion and lecture apparently interfered with learning the textbook, so here visual materials may have distracted the students from the verbal content upon which the tests were based. Such experiments make it clear that we still have little firm evidence upon which to base selection of visual materials for television or for other types of teaching. Selecting materials on the basis of "hunch" or of student reactions may be unwise, for such materials may lull the student into a comfortable feeling that he understands when in fact he doesn't.

The Pennsylvania State research does provide some support for the idea that television's effectiveness lies in giving the student a good view. In one experiment, students who had three weeks of instruction were given their choice of whether to finish the course in television classrooms or in the originating room. Depending upon the course, one-third to two-thirds of the students chose television; these students were predominantly those who had been assigned seats toward the back of the lecture hall.

Some theorists suggest that one important outcome of education is the tendency of students to identify with teachers. Proponents of television have suggested that television increases student identification with the instructor; thus it helps him feel closer to the instructor by improving his ability to observe facial expression, eye movement, and other individual characteristics not as clearly observable in a large lecture hall. While this may be true, the television students at Pennsylvania State reportedly did not often come in to talk to the instructor outside of class. Whether this would be paralleled by the behavior of students in other classes of the same size is not known. If individual contact with an instructor is important in any sort of learning, the large class, whether taught by television or not, may be ineffective.

Television compared with large and small classes. A second major project in closed-circuit instruction was that at Miami University (Macomber & Siegel, 1956, 1957a, 1957b, 1960). This group's research is of particular interest because it compared closed-circuit television with both large lecture classes and small semidiscussion classes, and studied the possible differential effect of different types of instruction upon students of differing abilities and attitudes.

The Miami University group attempted to use each method at its best. The television courses utilized professional directors, and audio-visual assistance was available in both television and other classes. The result was that the television classes gained and held student attention as well as most good classroom lectures. (But remember the New York University and Pennsylvania State findings that visually enriched production may produce inferior achievement.)

In the first experiment at Miami University, the main measure of achievement was the final examination in each course, and the television classes were not inferior on this criterion. In fact, in a human biology course the television students scored higher than the conventional classes, although other factors might have contributed to this difference. However, in the second year of telecasting, live teaching produced achievement superior to that in television classes in the second semester of the four courses, and in economics proved to be superior in producing gains in critical thinking. Results of third-year experiments were less unfavorable to television, although television classes in zoology proved inferior to conventional classes on a test of problem solving.

Since television is usually considered as a substitute for large classes, it is worth noting that large "live" classes did not consistently produce the inferior results of television instruction in cognitive outcomes but did tend to be inferior to small conventional "live" classes in effecting changes in attitudes.

Other experiments in college teaching by closed-circuit television. Among the other experiments on television teaching, those carried out at Purdue University, the State University of Iowa, New York University, and San Francisco State College were the largest and best designed. While these experiments have already been cited in our discussion of the Pennsylvania State and Miami experiments, a few points remain. One of these relates to teaching English composition by television. Even though the Purdue experiment (Seibert, 1958) used television for only two of the three instructional periods per week, and television students apparently had a good deal of practice in theme writing, television instruction proved to be significantly inferior to conventional instruction in several comparisons. The superiority of conventional teaching was most marked for students of lower ability. At New York University (Klapper, 1958) similar ability-level differences were not found, but there was evidence of superiority of conventional methods as measured by theme writing. Similar differences favoring conventional instruction using objective tests of achievement were found during the first semester but were reversed during the second semester.

At Purdue television instruction also was found to be inferior to conventional instruction in mechanical engineering (Seibert & Honig, 1959), military science (Kasten & Seibert, 1959), and, on some tests, in calculus (Seibert, 1957).

When a course demands the demonstration of small objects or parts, the use of television or films should be advantageous. In an experiment at Rensselaer Polytechnic Institute (Throop, Assini, & Boguslavsky, 1958) in teaching Strength of Materials, television was not inferior to conventional methods in teaching instrumentation and specimen behavior, but was inferior in teaching theory and familiarity with machinery.

Student ability generally does not make a difference in the relative effectiveness of television. At Miami University, low-ability students in Foundations of Human Behavior and Government achieved more in conventional classes than in television classes, but in physiology and zoology, the opposite was true. However, student ratings of television instruction are inversely correlated with student ability. While the best television instructors are liked by all types of students, the better students ordinarily dislike television and large classes more than do the poorer students. Students feel that television hurts both the instructor's pacing of the instruction and his stimulation of interest. Attitudes toward television, however, do not greatly affect achievement. Students who dislike television achieve almost as much (or little?) as those who like it, and, interestingly, such attitudes tend to reflect those of the proctors in the viewing rooms.

In both large lectures and in television sections, students complain of lack of contact with the instructor, but Miami students disliked television less than large lectures, while New York University students tended to prefer lectures. The attitudes toward television of both groups tended to become more negative during a second semester of television. If they could have the same instructor, students generally preferred a small section to television or a large class. But they preferred television or a large class to a small class if they could be sure of having an excellent instructor in the television or large class but had to take their chances in electing a small class. This is probably a fairly realistic alternative. The student's choice may not always be wise, however, for in experimental comparisons Miami graduate assistants proved to be inferior (as measured by student achievement) to regular staff members in one course and superior in another. Moreover, students taught by graduate assistants did more outside reading.

Uses of television. From our hypothesis that television would be of most value in courses depending upon visual presentation of information, we might expect it to be more effective in science and engineering courses than in social sciences and humanities courses. From our second hypothesis that television would be of less value in classes where interaction between students and instructor is important, we might expect it to be relatively less effective in psychology, speech, and language courses than in courses usually taught by lecture. However, such comparisons are difficult to make. As we have seen, students learned as much by television as in conventional classes in chemistry at Pennsylvania State and in engineering at Purdue (Seibert & Honig, 1959), but in general psychology the television students both at Purdue and Pennsylvania State did not do as well as students in conventional classes. This evidence is in line with our hypothesis but certainly not conclusive.[8]

It was suggested above that television should be especially effective in courses requiring the use of demonstrations. If we assume that science courses are more likely to use demonstrations than are other courses, we can examine the experimental results to determine the relative effectiveness of television instruction in science and in other courses. Such an appraisal provides some support for our hypothesis, for if we simply look at the direction of the differences, about half the experiments in science classes favor television and half favor conventional instruction. In nonscience courses well over two-thirds of the differences favor conventional teaching. Similarly, the one experiment in which television was significantly superior to conventional instruction was in a biology course, while six of seven experiments in which conventional teaching proved superior were in nonscience courses.[9]

Whether television or film should be used in a course requiring demonstrations of small phenomena seems to be largely a matter of cost. However, there seems to be a tendency to forget the possibility of reproducing diagrams and photographs as an alternative to television. While television and film seem to be clearly advantageous as methods of presenting relations between moving parts, many of the "visuals" used on television do not involve movement, and many photographs, enlargements, slides, and lithoprints can be made for the price of one television camera or film.

Insofar as student-instructor interaction is important in teaching a course, television appears to be of little help. Television does not permit more students to talk, even with two-way audio connections. The student's opportunity to participate is an inverse function of the number of students in the class. If actual participation is important, larger classes should be less effective whether they are taught in one classroom or by television. Television may permit students to gather in smaller groups (as in the Iowa experiments) but the Iowa results (Stuit, et al., 1956) suggest that students feel less free to participate in television classes than in regular classes. The 90-minute classes at Miami used television during the first part of the period to provide the stimulus for discussion; this procedure seems to be successful. However, the Pennsylvania State studies did not reveal any gains from discussion or two-way communication.

---

[8] Few experiments have adequately evaluated the possible assets of television as a visual medium. Most achievement tests are verbal. We would expect television to reveal its unique values when outcome measures require visual recognition or discrimination. While the objectives of most college courses emphasize abstraction and verbalization, there must be some areas in which visual relationships are important.

[9] The preceding comments have been primarily concerned with closed-circuit television instruction on campus. While it is difficult to make adequately controlled studies of the educational effectiveness of broadcast television, it seems quite clear that students motivated enough to take a television course for credit at home learn well and have favorable attitudes toward television (Dreher & Beatty, 1958; Evans, Roney, & McAdams, 1954; Lepore & Wilson, 1958).

Cutler, McKeachie, and McNeil (1958) compared the effectiveness of teaching equal-sized groups in face-to-face groups and by *telephone*. Both groups showed significant learning and attitude change, and there was no significant difference in the effectiveness of the two methods. Thus, if economy in instruction is desired, perhaps the expense of television cameras and receiving tubes is unnecessary. One of the few experiments comparing the effectiveness of radio and television, however, showed better learning and retention for television (Paul & Ogilvie, 1955), and experiments conducted a generation ago tended to find differences favoring face-to-face instruction over radio or printing (e.g., Wilke, 1934).

Television can, however, have certain *advantages* in promoting interaction. On any large college campus, one difficulty in education outside the classroom is that students have few common educational experiences. Since their common experiences tend to be social or athletic, these are likely to be the usual topics of conversation. Stephens College (1955) met this problem by developing a required course, "Ideas and Living Today," which consisted of brief television lectures viewed by students in small faculty-led discussion groups. The course is scheduled just before lunch so that discussions spill over into the dining halls. The course thus provides a common intellectual experience which can stimulate interstudent education outside the classroom—an element of education lacking in many colleges today.

Even if closed-circuit television were a potential answer to all our educational problems, its success would probably depend to some extent on faculty acceptance. Apparently all colleges using closed-circuit television have found that many faculty members are skeptical of its value. Polls generally indicate that faculty members are willing to teach by television if necessary but prefer traditional teaching methods (e.g., Carpenter & Greenhill, 1955).

Faculty acceptance of television seems to be greatest when the faculty has participated in planning for it. For example, the faculty of Stephens College seems to have accepted television more fully than the faculties of other institutions, and this is probably due in no small part to their involvement in planning for their television course.

A second factor in the acceptance of television at Stephens is that the course offered does not set a particular department or instructor apart from the rest of the faculty. The course offered at Stephens is interdivisional and involves almost all of the faculty. Since the course is a new one, no local professor was pushed aside to make room for a "Master Teacher."

When television enables a department to provide instruction which would otherwise be difficult or impossible, acceptance also seems to be good. Thus Pennsylvania State found no faculty objections to teaching electrical engineering by television when a large enrollment plus a staff shortage resulted in several unstaffed sections.

Finally, the support of administrative officials is also of obvious importance. Since television teaching is difficult and time-consuming, faculty members are much more likely to participate if they feel that this activity is valued not only by their colleagues but also by their administrative officials. (One evidence of such support is a reduction in teaching load for television instructors.)

Taking the results of all research on television instruction, we feel safe in concluding that television instruction for a complete course is inferior to classroom lectures in communicating information, developing critical thinking, changing attitudes, and arousing interest in a subject but that this inferiority is probably not great. This conclusion may surprise the reader who has seen publicity releases reporting no significant differences between classroom and television instruction. In our consideration of methodological issues, we noted the logical fallacy in concluding that no difference exists because the difference found was not great enough to disprove the null hypothesis. This is particularly pertinent here because a review of a number of studies of television leaves one with quite a different conclusion than he might draw from a review of just one. In 20 of the 26 well-controlled experiments reviewed, conventional classes were superior to television classes in achievement. Although few of the differences were statistically significant by themselves, simple application of the sign-test indicates that the differences were not random. In contrast with research comparing other instructional methods, the consistency of results favoring conventional instruction over television is unusual.[10] When one weighs heavily the necessity for accommodating higher education to large numbers of students, however, the differences between television and conventional instruction seem very small. It may be that researchers are reluctant to report findings contrary to the hypotheses of the foundations that support them, and their reports are thus as gentle as possible.

Films

The great mass of research on instructional films is relevant to our topic, even though most of it is not concerned with college teaching. While we cannot summarize all of the relevant studies, certain principles have emerged. (For a more complete analysis, see Chapter 12 and Miller, 1957.)

1. Students can learn from films and usually do learn at least as much as from a poor teacher (VanderMeer, 1950).

2. Such learning is not confined to details but may include concepts and attitudes (Hoban & Van Ormer, 1950; Kishler, 1950; Mertens, 1951).

3. Repeating the film increases learning (McTavish, 1949).

4. Students learn how to learn from films; i.e., students with previous experience in learning from films learn more than those with no previous experience with instructional films (VanderMeer, 1951).

5. Presenting pictures is more effective than presenting words as stimuli in rote association tasks such as learning a foreign language (Kopstein & Roshal, 1954; May & Lumsdaine, 1958).

6. Participation increases learning (Hovland, Lumsdaine, & Sheffield, 1949).

Much of the preceding discussion of instructional television is relevant to teaching by films. The chief differences seem to be that the movie camera has greater mobility than the television camera, it is more expensive to produce a film than a television lesson (assuming that the television equipment is already available), and "live" television is presumed to have greater immediacy. Unfortunately, we have little evidence on the educational importance of immediacy. Intuitively it seems that students would feel more involvement in watching a television professor on their own college's staff, knowing that they are seeing him at the actual moment of performance, than they would feel in watching a film made at some earlier time. Whether or not such differences in involvement do occur (and if they occur, whether they make any difference educationally) is still unknown.

Tests and Teaching Machines

While we usually think of testing procedures in terms of their validity as measures of student achievement, their function as instruments for promoting learning may be even more important. After dismal recitals of nonsignificant differences between differing teaching methods, it is refreshing to find positive results from variations in testing procedures.

In some of the earliest experiments in this area, Jones (1923) found that immediate testing after a psychology lecture resulted in improved retention. The good effects of testing persisted or increased over an eight-week period.

---

[10] Note that most research has dealt with television as a substitute for conventional instruction. The potential of television as a tool for enriching classroom teaching has not been assessed.

Jones's results supporting the value of immediate feedback are supported by an experiment in a government class (Fitch, Drucker, & Norton, 1951) in which students having weekly noncredit quizzes made better scores on monthly tests than a nonquizzed control group. Similarly, in a remedial English course at Purdue (Maize, 1954), students who wrote 40 themes evaluated in class made greater improvement on a test of English usage than a group which had workbook drill and wrote 14 themes, individually corrected by the instructor. Similar results were noted in the study by Guetzkow, Kelly, and McKeachie (1954). May and Lumsdaine (1958) reported that learning from film is also positively influenced by participation and feedback devices.

Tests provide knowledge of results, one of the major elements in learning, and we would expect that the more information contained in the feedback, the greater its value. In an experiment in the Air Force (Stone, 1955), performances benefited from return of multiple-choice tests together with information about why the alternative chosen was wrong and why the correct alternative was right. This technique proved superior to four other techniques which gave less complete knowledge of results ranging down to returning only a score on the total examination. Similar results were obtained by Bryan and Rigney (1956). In a related study, McKeachie and Hiler (1954) found that students required to answer study questions performed better on test questions in the same area than students not given the study questions and tended to do better than those whose answers were not required or graded. Thus the simple principle that knowledge of results facilitates learning is one of the few generalizations clearly supported by research on college teaching.

Until this point, the reader of this section may wonder why testing procedures have been classified under the general heading, "Automated Techniques." The bridge to that general topic is found in the work of Pressey (1926, 1950; also included in Lumsdaine & Glaser, 1960), who published the results of an extensive program of research with tests which students scored for themselves by punching alternatives until they hit the correct one. Four types of studies were carried out, and Pressey concluded that (1) the self-scoring characteristics represent a tremendous saving in time; (2) test taking is transformed into self-instruction by the immediate knowledge of mistakes; (3) supplemental use of the tests improves performance on regular objective tests; and (4) even more automatic self-scoring devices can be devised. This last conclusion proved to be prophetic, for there are now many models of "teaching machines."

The term *teaching machine* now refers to a device for presenting content and questions in predetermined sequences and providing immediate knowledge of results to an active learner. Teaching machines permit the learner to proceed at his own rate. The successive questions proceed in tiny steps from the simple to the complex. If a student makes a series of correct responses, he may adjust the machine to skip some steps. If he fails items, they are repeated. The program of the lesson may include hints or other guidance. Thus teaching machines or workbooks have many theoretical advantages over lecturing or other conventional methods of instruction. Unfortunately, little experimental work has been done at the college level to determine the limits of their usefulness. At present it appears that learning is enhanced by programed presentation but that programed presentations require more time than conventional textbooks (Angell, 1949; Evans, Glaser, & Homme, 1960; Little, 1934; Stephens, 1953; Stolurow, 1960). At present, much developmental work is being directed not only at the basic problem of devising appropriate sets of items and determining optimal modes of presentation but also at development of machines flexible enough to adjust to individual differences in the background and ability of the learners. (See Galanter, 1959; Lumsdaine & Glaser, 1960; and Chapter 12 in this Handbook.)

As we pointed out earlier, one advantage of the skilled teacher is his ability to utilize feedback in adapting his teaching to individual and group differences. Moreover, the instructor often provides feedback which not only tells the learner that he is wrong but tells him why he is wrong. Experiments at the System Development Corporation (Coulson, 1959; Coulson & Silberman, 1959) suggested that teaching incorporating such human characteristics is more effective than teaching by the typical fixed-sequence machines. (In these experiments, instead of using teaching machines to simulate human teachers, the experiments used humans to simulate teaching machines.)

Printing

Books used as substitutes for the live teacher have been in existence for some time. Unlike other technological aids, printing has found wide acceptance by college faculties. Perhaps a few centuries of use will produce the same degree of acceptance of television and teaching machines.

Despite the age of printing as a technique, there is relatively little research on its use, although much of the research on independent study, of course, bears upon students' ability to learn from books. In an early study, Greene (1928) found that students learned as well from reading material themselves as from listening to the same material read aloud. The better students, however, profited more from reading than from listening. Corey (1934) found better immediate recall for reading than for a lecture, although the difference disappeared on a later test. Oral presentations, however, seem to be superior to printing in changing attitudes or opinion (Knower, 1935, 1936; Wilke, 1934). Cantril and Allport (1935) suggested that critical attitudes are favored by reading.

Becker and Dallinger (1960) compared three teaching methods in a course on Communication Skills. One method reduced class meetings from four to three a week and presented basic principles through assigned and optional readings. A second method used kinescope presentation of basic principles by experts, supplemented by discussion led by the classroom instructor. The third method, like the second, used four class periods a week for normal instructional procedures. An analysis of variance of achievement tests revealed significant differences between instructors but not between methods. Both students and instructors, however, preferred the normal method.

One recent development is the "programed textbook," an instructional book developed by utilizing the learning-in-small-steps sequence of the conventional teaching machine (Glaser & Homme, 1958). It seems entirely reasonable that books designed specifically to achieve certain objectives could be more effective than our present textbooks. Unfortunately, we are probably better able to devise sequences optimal for cognitive learning than for arousing and maintaining motivation. Information obtained by research on teaching-machine programs should be valuable to the textbook writer.

A recent study on the learning of college text-type materials by Ausubel, Robbins, and Blake (1957) probably will not seem significant to those who are unfamiliar with classical learning theory. Psychologists have used the concept of "retroactive inhibition" to describe the fact that in learning experiments new learning tends to interfere with memory of previous learning. Recent research (Underwood, 1957) has cast some doubt on the interpretations of this phenomenon. The experiment by Ausubel and his associates demonstrated that memory of a passage on Buddhism was not affected adversely by later study of a passage on Christianity. In fact, when passages repeated points from the earlier material, memory was actually facilitated. One critical question for the instructor, however, is still unanswered: What is the effect of giving one point of view in a lecture and assigning opposing points of view in readings? One would guess that this technique would be ineffective as measured by a test of retention but might contribute positively to ability to think creatively about a problem.

The research to date indicates that television, films, teaching machines, and books can be used to achieve educational objectives. Their usefulness depends on the objective, characteristics of the students, and the excellence of the materials. Research at present reveals no likelihood that these devices will eliminate the need for face-to-face contacts between professors and students.

STUDENT CHARACTERISTICS RELATED TO EFFECTIVE TEACHING

Most teachers are aware that differences between students are not taken into account sufficiently by our usual teaching methods. Experienced teachers have felt for years that no one teaching method succeeds with all kinds of students. One reason for the host of experimental comparisons resulting in nonsignificant differences may be simply that methods optimal for some students are detrimental to the achievement of others. When mean scores are compared, one method thus seems to be no different in its effect from any other.

It is encouraging to note that when an instructor is aware of the nature of differences among his students, his teaching is more effective. In an experiment in which teachers of physics were given increased knowledge of their students' personal backgrounds for one of their two classes, the students made significantly greater gains in achievement and rated their instructors as being more effective in the experimental (informed) classes (Sturgis, 1958).

We have already noted some analyses of teaching methods taking such individual differences into account. For example, using analysis of variance, Guetzkow, Kelly, and McKeachie (1954) found that students differing in intelligence or in preferences for teaching methods were not differently affected by the three methods used in their study. Hudelson (1928) found similar results in his studies of class size. Macomber and Siegel's results (1956), although not strongly opposite, did reveal a tendency for high-ability students to gain more in course-related attitudes in small rather than large sections. They also report a small superiority in achievement in television or large classes for those who initially held favorable attitudes toward the method used. Other personality measures did not, however, predict differential achievement in large and small classes. Ward's (1956) results also suggest that the ablest students benefit most from small groups. Comparing group study and lecture-demonstration methods in a physical science course, he found that the group method resulted in better achievement on a measure of understanding and problem solving for the abler students. The poorer students, however, benefited more from lecture-demonstration.

Parsons, Ketcham, and Beach [11] (1960) reported that the abler students are less satisfied the less their responsibility for setting goals of learning. Calvin, Hoffman, and Harden (1957) found in three experiments that less intelligent students consistently did better in group problem-solving situations conducted in an authoritarian manner than in groups conducted in a permissive manner. The same difference did not occur for bright students. The experimenters suggest in a footnote that the inferior performance of dull students in permissive groups may be due to an inferior ability to adjust to a change from the usual methods of conducting groups. If this were true, we should expect brighter students to adjust more easily to new methods than less intelligent students. So far, this has not been demonstrated, although the trends reported seem to be in that direction. For example, in the Miami experiment, high-ability students did as well in television classes as in conventional classes although they had more negative attitudes toward television than the less able students had. On the whole, our best guess would be that face-to-face discussion teaching is more advantageous for the more intelligent students than for those with less ability.

---

[11] T. S. Parsons, W. A. Ketcham, & L. R. Beach. Students, teachers, and fellow students: A study of instructional interactions and outcomes. Unpublished paper, 1960. (Dittoed)

The analysis of student characteristics in studies such as those above is worth while, but, except for the study by Calvin, Hoffman, & Harden (1957), analysis of individual differences was largely peripheral to the main purpose of the studies and was undertaken with little theoretical basis to guide it. As focus shifts to student characteristics, research designs should employ teaching method variables and dimensions of individual differences bearing some theoretical relationship to one another, as suggested by Cronbach (1957). For example, if we are going to investigate intelligence as a student characteristic affecting learning under different teaching methods, we need to develop a more specific theory about the conditions which are effective in producing learning for students of high and low intelligence.

Independence

A pioneering step in the direction of relating personality and response to teaching methods was taken by Wispé (1951), who used TAT-like measures to differentiate three types of students: the "personality-insecure" student (51 per cent of the sample), the "satisfied" student (26 per cent of the sample), and the independent student (23 per cent of the sample). The "insecure" student had unfavorable attitudes toward permissive teaching. The "satisfied" student had favorable attitudes toward instructors, fellow students, and both directive and permissive teaching methods. The "independent" student was highly verbal and wanted more permissive teaching no matter what method his instructor used. The independent student had moderately favorable attitudes toward fellow students and instructors but was likely to direct aggression against the instructor in directive classes.

In conjunction with Wispé's results, a study of factors associated with student choice of directive or nondirective classes is of interest. In their study of Springfield College students, Ashmus and Haigh (1952) found, unlike Wispé, that almost equal numbers chose each method. Students choosing nondirective classes did not differ in intelligence or grades from students choosing directive classes, but they were more likely to have had previous experience in nondirective groups and tended to be more flexible, to be better able to cope with ambiguity, and to have more self-insight (Haigh & Schmidt, 1956).

A cluster of variables probably related to Wispé's is Patton's (1955) "acceptance of responsibility for learning." Patton found that the degree to which the student accepted responsibility was positively correlated with gain in psychological knowledge, gain in ability to apply psychology, rating of the value of the course, and interest in psychology.

But what sort of student will accept responsibility in such a course? Patton (1955) found that the students who liked his experimental class and assumed responsibility were likely to be independent of traditional authority figures and high in need for achievement. Koenig and McKeachie (1959) also found that women high in need for achievement preferred independent study to lectures. Something similar may be involved in the suggestion in the Oberlin studies (McCollough & Van Atta, 1958) that students who are less rigid and less in need of social support are likely to profit more from independent study than those students who are not as independent.

Despite the variety of measures used, the studies cited in this section show some consistency in finding that a certain type of student, characterized as independent, flexible, or in high need for achievement, likes and does well in classroom situations which give students opportunity for self-direction.

Anxiety and Sex Differences

One of the most active intersections of experimental psychology and experimental study of personality has been the area involving the effect of anxiety upon learning. The impetus for much of this work came from deductions from Hullian theory by Spence and Taylor (1951). Essentially, this theory suggests that a high level of anxiety facilitates simple learning, but beyond an optimal point hampers complex learning. Since academic situations sometimes produce high levels of anxiety, we might expect damaging effects. Beam (1955) showed that doctoral examinations and oral reports did decrease ability to learn nonsense syllables but improved conditionability.

Since individuals obviously differ in their reactions to stress, a number of investigators have attempted to devise questionnaires which might differentiate individuals who respond differently in stressful situations. In addition to the Taylor Manifest Anxiety Scale used in the studies cited above, a number of experimenters have used the Mandler-Sarason (1952) questionnaires. Individuals who scored as highly anxious on this questionnaire responded with poorer performance after reports on their success or failure while those with low anxiety improved.

The Interfering Tendency Questionnaire (which correlates well with the Taylor scale) developed by Waterhouse and Child (1953), differentiates individuals responding differently to failure. Under neutral conditions, individuals scoring high in interfering tendencies do better than low scorers, but after failure the high scorers do worse. This result was confirmed by Williams (1955), who also found that reaction to different types of failure is affected by achievement motivation.

One of the most comprehensive studies of student personality was carried out by Gaier (1952). Using the Rorschach, Gaier determined the relationship between certain personality characteristics and students' thoughts in small discussion groups. He found that anxiety readiness, as measured by the Rorschach, correlated negatively with scores on a reading test and a test requiring comparisons of familiar and unfamilar materials. Rigidity predisposition correlated positively with class time spent in negative thoughts about other people and negative thoughts about ideas expressed in class.

Since anxiety is generally believed to be increased by uncertainty, we would expect the anxious person to work most effectively in a highly structured situation. This hypothesis is partially supported by the research of D. E. P. Smith and his co-workers (1956), who found that anxious students who were permeable (sensitive to stimuli, impulsive, socially oriented, and low in ego strength) made optimal progress in a remedial reading course when taught by directive methods. Impermeable anxious students, however, were unaffected by differences in teaching methods. H. C. Smith (1955), on the other hand, found that students with high anxiety and low initial achievement gained more on achievement tests and were more highly satisfied in a "teamwork" class than in a conventional lecture.

The relationship between anxiety and performance on classroom examinations administered under varying conditions has been the subject of several experiments. To test experimentally whether the aggression created by tests might be dissipated by permitting students to write comments on tests, half of the students in a University of Michigan experiment were given answer-sheets with spaces for comments and half were given standard answer-sheets. Measures of students' feelings about the tests failed to show any difference between the two groups; but the students who had the opportunity to write comments made higher scores on the test. These results held up in a series of experiments (McKeachie, Pollie, & Speisman, 1955).

It thus appeared that some students were too highly motivated—at least for performance on tests. The findings indicated that student anxiety during classroom examinations builds up to such a point that it interferes with memory and problem solving. Reducing the stress of the examination by permitting students to write comments resulted in improved performance.

This interpretation was supported by Calvin, McGuigan, and Sullivan (1957), who found that students who were given a chance to write comments on an achievement test were superior to control students in their performance on the second half of the test and that the students who made the greatest gain were the highly anxious students as measured by the Taylor Manifest Anxiety Scale. Similarly, W. F. Smith and Rockett (1958) found that instructions to write comments significantly interacted with anxiety, helping the performance of high-anxiety students but hurting the performance of students low in anxiety.

Carrier (1957) investigated the manner in which individual differences in four personality variables affected performance in more and less stressful testing situations. In the "low-stress" condition, students had answer-sheets with spaces for comments, and the test was administered in a friendly, relaxed fashion. In the "high-stress" condition, the test was taken in a strange room, was administered by an austere stranger who announced that he would answer no questions during the test, and students were told that a point would be deducted for every wrong answer but that they must answer every question.

Carrier found that one of the most important variables determining reaction in his experiment was sex. Women were much more detrimentally affected than men by his stress situation. Individual differences in need for achievement and need for affiliation failed to predict reactions to the stress situation. However, the permeable students, as measured by Smith's SA-S scale (Smith, et al., 1956), were detrimentally affected by stress.

In a later experiment (McKeachie, 1958c), half of a group of 350 students taking a test received conventional IBM answer-sheets; half received answer-sheets with space for comments. In addition, half of the students received a tranquilizing drug, meprobamate ("Miltown"), from a physician associated with the Institute of Mental Health Research, while the other half received a placebo. All of the students had taken a number of tests of anxiety and of need for achievement a week before the examination. If students tend to be too anxious, such a drug should improve test scores. The results did not confirm this hypothesis. Students who actually had the Miltown reported experiencing less anxiety during the examination than did the placebo group, but they did not make better scores.

The really interesting result of the experiment was the interactions. The sex-drug interaction was significant both in its effect on performance and on anxiety, with women benefiting from the drug more than men. Thus sex once again turned out to be an important variable. The results make sense if we assume a curvilinear relationship between anxiety and performance, with women more anxious than men. Thus, reduced anxiety should result in improved performance for women, but poorer performance for men.

An Air Force experiment on teacher-student interviews also revealed complex interactions. Although the interviews were not effective overall in influencing achievement, Hoehn and Saltz (1956) found that anxious students tended to be helped by interviews while rigid students were more likely to fail if interviewed. The results were further affected by type of interview; interviews in which students were encouraged to gripe produced the interaction noted above while interviews oriented toward the student's goals and sources of satisfaction did not. If this result can be generalized to the testing situation, we might expect the "comments" technique to be unsuccessful for rigid students.

The experimental results on interaction of anxiety and teaching variables are tantalizing enough to stimulate further work, but they are not consistent enough to lead to any stable generalizations. The fact that differences in sex and in other personality variables, such as permeability, interacted with anxiety in the experiments above suggests that multivariate designs are necessary to explore this area more fully.

Authoritarianism

One of the most intensively studied research variables of the past decade has been "authoritarianism." It is not surprising, therefore, that this variable has been studied in relation to college learning.

Watson (1956) studied the effect of permissive and restrictive teaching and testing methods upon students differing in authoritarianism and permeability (extraversion). The methods were not differentially effective as measured by achievement tests but student satisfaction was affected by testing methods. Highest satisfaction resulted when one was tested in an atmosphere appropriate for his needs, i.e., permissive for permeable, restrictive for impermeable. This finding is in line with the finding of Bendig and Hountras (1959) that authoritarian students prefer a high degree of departmental control of instruction and with the Hoehn-Saltz results reported earlier. Somewhat more complicated is the finding of Koenig and McKeachie (1959) that flexible nonauthoritarian women participate more in small groups than rigid women. For males, however, the relationship is reversed; flexible men are less likely to participate than rigid men.

Beach (1960) studied the personality variable of sociability as a predictor of achievement in lecture and small-group teaching methods. In the lecture section the nonsociable students (as measured by the Guilford Inventory of Factors STDCR) achieved significantly more than the sociable students; in the small-group sections the results were reversed.

By far the most extensive studies of personality characteristics related to college achievement are those of Stern, Stein, and Bloom (1956) and Pace and Stern (1958). These authors have developed not only a variety of methods of assessing student personality, but also a scale for assessing the pressures on students in various college environments. First, they set up a theoretical model of the personality syndrome being studied (a sort of rigid, uncritical, authoritarian pattern), created the Inventory of Beliefs as the chief predictive instrument for the theoretical model, and then obtained extensive data over a period of one year on an incoming group at the University of Chicago. In addition, after the initial data had been analyzed, an individual was selected who seemed closest to the syndrome described, and a case study was made. Correlations were found between the personality syndrome taken from the extremes of the Inventory scores and (1) excellence in placement tests, (2) grades in subsequent courses, (3) ability to adjust to university life at Chicago, and (4) classroom behavior.

Extending the personality model to include major needs, Stern, Stein, and Bloom created the Activities Index as a measure of these needs and have used it and the Inventory of Beliefs in a number of studies in academic settings.

For example, as we saw earlier, students high on a variable akin to authoritarianism were found to gain more when taught in a homogeneous group. The instructor who taught this section found that he had to resist pressures from the students for lectures. He secured his good results by using many direct questions, encouraging student responses, and by vigorously defending absurd positions which even authoritarian students would argue against (Stern, 1962).

In addition, Pace and Stern have studied the relationship between student needs and the forces affecting students in differing colleges. Their College Characteristics Index and the Activities Index reveal great differences between colleges in such characteristics as the tendency to support the need for order or the need for achievement. Presumably, the effects of variations in teaching methods not only interact with student characteristics but also with aspects of the college's culture. Thus, apparently inconsistent results may simply reflect differences in the cultures in which the experiments have been conducted.

In summing up the studies on the interaction of personality characteristics and teaching methods as that interaction affects student learning, it is safe to say that no major break-through has occurred. The results so far do, however, appear promising. Fortunately, multivariate statistical techniques now permit precise analysis of the sorts of complex interaction involved in teaching. Although we have little data on this point (Guetzkow, Kelly, & McKeachie, 1954; Jenkins, 1952), it is probable that instructor characteristics and content characteristics also interact with teaching methods. For example, in both the Michigan State and Army studies of educational television, some instructors proved to be more effective on television than face to face while others proved to be more effective in face-to-face teaching. Similarly, the experiments of Wakely and his associates (1960) and Russell (1951) indicate that some teachers are more effective with the abler students, while others are more effective in teaching the poorer students. Such interactions pose exciting problems for investigation.

THE ROLE OF FACULTY ATTITUDES

What can we say about the work of the college teacher? Clearly it is not possible to detail in a few summary statements the "best" methods of teaching. Nevertheless, a conclusion that it does not make any difference which methods are used is clearly unjustified. Rather, recent research suggests that decisions about teaching methods do have important consequences in terms of differential achievement of the differing objectives of a course, differential effects upon different types of students, and probable differential effects depending upon other factors such as the instructor, the course content, and the over-all "climate" of the institution. To analyze such complexities would obviously be a task for an electronic computer. In the absence of the data necessary for such an analysis, we must, as in other frontier areas, depend upon expert judgment. Most of the reports of research on teaching neglect to report the reactions of the faculty involved (except in the case of television, where they are generally negative). Yet until we gain more confidence in our evaluation tools, we are almost forced to weigh faculty judgment heavily. This writer would argue that faculty judgments of teaching methods are extremely important for the educational researcher even aside from their possible validity.

As was pointed out earlier, we seldom know how well a particular method was used in experimental studies of teaching methods, but it seems very likely that the effectiveness of a method depends upon the competence and enthusiasm of the teacher. If the teacher is important, his enjoyment of the method becomes a critical variable. Thistlethwaite (1960) reported that National Merit Scholars consider one of the critical variables influencing their choice of a field to be the instructor's enthusiasm. Such enthusiasm is not likely to be communicated if the instructor finds teaching distasteful. Thus, even though we found that a particular method when ideally used is superior to other methods, we should be dubious about urging its widespread adoption if teachers using it become bored or dissatisfied.

We have already discussed the probability that teaching effectiveness is not a direct function of the method used but rather of the interaction of teacher and method. One instructor may be an excellent lecturer, another a fine discussion leader. But so far our researchers have been unable to obtain large enough samples of instructors to be able to generalize to populations of instructors. It may be, for example, that, even though the discussion method seems potentially superior to the lecture method in achieving higher-level educational outcomes, lecturing is more effective for inexperienced teachers. Perhaps if large numbers of inexperienced teachers are to be employed, lecturing should be the method chosen.

Some administrators and foundation executives seem to be annoyed by the stubbornness of some professors in not accepting larger classes or teaching by television. But if we look at the rewards in teaching, the attitudes of faculty members toward class and institutional size become more understandable.

What are the satisfactions in teaching? Certainly one is the pleasure of seeing a student develop. Another is the pleasure of intellectual interchange with young people possessing questioning minds and fresh ideas. Perhaps a less laudable but nonetheless real satisfaction is that found in having disciples who respect and admire us. These satisfactions are difficult to secure without close, sustained personal contact with students. If we are to know students well enough to see their progress, small classes are important. Such classes permit more individual interaction with students and permit the instructor to use term papers, essay tests, and other evaluation methods which give him a greater understanding of what the student is thinking. While objective tests might also be used to give a personalized understanding of each student's strength and weaknesses, few teachers take the time to analyze the patterns of right and wrong answers that would be necessary for such understanding. Such an analysis becomes prohibitively time-consuming in a large class, whatever the testing methods used.

Many professors conscientiously attend some student teas or other social functions to promote contacts between students and faculty. In small colleges it is likely that the professor will meet at such functions some of the students he has taught and will have an opportunity to use this contact to gain greater understanding of the students and perhaps even to stimulate their thinking. But the larger the college, the less the statistical chance that he will meet students he knows at such functions; in fact, the larger the college, the less chance that he will ever meet again the student he meets at a tea. This means that even the professor who conscientiously devotes a portion of his time to such "informal" contacts with students is unlikely to have significant encounters with them. Because the professor is nearly always dealing with strangers or near-strangers, the intellectual interchange is almost perforce limited to polite inquiry about the student's academic and vocational aspirations or conversation about the current film at the campus theater. In short, any satisfaction received from observing and contributing to a student's growth must ordinarily come during the semester (or at most two semesters) that the student is enrolled in one's class. No matter how powerful one's impact, it is asking a great deal to expect it to have great, noticeable effects in 16 weeks.

Size of an educational institution has a very similar relationship to the quality of education students receive from one another. The large institution with a student body of heterogeneous background offers students an opportunity to gain breadth, tolerance, and new perspectives from their contacts with one another. But large size is likely to reduce educational values by reducing intellectual interchange between students. There is certainly no reason why a student at a large college could not discuss with another student an interesting problem raised by one of his professors. But he is probably more likely to do so if he is living near another student who is also familiar with the problem and concerned about it. In a large college the statistical chances that another student in the same class will be in the same living group are smaller than in a small college. Students in a large college with many courses, and even many sections of the same course, have few common intellectual experiences. Consequently it is difficult for them to communicate about intellectual problems outside of class, and the common concerns which become the basis of social communication are football, the student newspaper, dating, and the dormitory food. With such barriers to interstudent education, the professor misses the good feeling one experiences when he finds that his teaching has provided an intellectual stimulus reaching beyond his classroom.

Of course there are also satisfactions in teaching by television or in a large class. One can gain a very satisfying sense of power from knowing that one is communicating one's ideas to a large number of students. The roar of laughter at a joke well told is music to a lecturer's ear. The satisfaction of carrying through without interruption a well-planned lesson is satisfying to the "Master Teacher" whose performance is televised.

Although these are valid satisfactions, they seem less directly related to the goals of education than the satisfactions associated with observing student development. What would a college be like if its faculty were made up largely of teachers whose satisfactions were received primarily from being a good performer?

As colleges increase in size in order to cope with a growing student population, there is a natural tendency to routinize and automate educational processes in the interest of increased efficiency. In industry, assembly line methods have long been effective. Yet, in recent years, industry has found that workers are even more efficient if, instead of performing one specific, repetitive task, their jobs are enlarged enough to provide variety and interest (e.g., Mann & Hoffman, 1960). While there is little likelihood that college administrators will intentionally insist upon uniform teaching methods, increasing class size indirectly limits the professor's choice of teaching methods, reducing his ability to select the methods best suited for his objectives and reducing his satisfaction in teaching.

Enjoyment of teaching is important not only for the enthusiasm which the professor communicates to his students but also for his interest in continued improvement. These important values are likely to be lost if teaching becomes so routinized and depersonalized that it is no longer fun. The motivated teacher can respond to feedback from his students so as to achieve better and better approximations to optimal solutions to the problems of teaching. As additional information from research accumulates, as better conceptualizations emerge, he should be able to do an even better job.

REFERENCES

Adams, J. C., Carpenter, C. R., & Smith, Dorothy R. (Eds.) *College teaching by television.* Washington, D.C.: American Council on Education, 1958.

Anderson, R. C. Learning in discussion: A resumé of the authoritarian-democratic studies. *Harvard educ. Rev.*, 1959, 29, 201–215.

Anderson, R. P., & Kell, B. L. Student attitudes about participation in classroom groups. *J. educ. Res.*, 1954, 48, 255–267.

Angell, G. W. Effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. *J. educ. Res.*, 1949, 42, 391–394.

Asch, M. J. Nondirective teaching in psychology. *Psychol. Monogr.*, 1951, 65, No. 4 (Whole No. 321).

Ashmus, Mable, & Haigh, G. Some factors which may be associated with student choice between directive and nondirective classes. *Amer. Psychologist*, 1952, 7, 247. (Abstract)

Atkinson, J. W., & Litwin, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *J. abnorm. soc. Psychol.*, 1960, 60, 52–63.

Ausubel, D. P., Robbins, Lillian C., & Blake, E., Jr. Retroactive inhibition and facilitation in the learning of school materials. *J. educ. Psychol.*, 1957, 48, 334–343.

Axelrod, J. Group dynamics, nondirective therapy, and college training. *J. higher Educ.*, 1955, 26, 200–207.

Bainter, Monica E. A study of the outcomes of two types of laboratory techniques used in a course in general college physics for students planning to be teachers in the elementary grades. Unpublished doctoral dissertation, Univer. of Wisconsin, 1955. (*Dissertation Abstr.*, 1955, 15, 2485–2486)

Balcziak, L. W. The role of the laboratory and demonstration in college physical science in achieving the objectives of general education. Unpublished doctoral dissertation, Univer. of Minnesota, 1953. (*Dissertation Abstr.*, 1954, 14, 502–503)

Bane, C. L. The lecture vs. the class-discussion method of college teaching. *Sch. & Soc.*, 1925, 21, 300–302.

Barnard, J. D. The lecture-demonstration versus the problem-solving method of teaching a college science course. *Science Educ.*, 1942, 26, 121–132.

Baskin, S. Experiment in French language instruction. *Antioch Coll. Rept.*, Oct., 1960.

Beach, L. R. Sociability and academic achievement in various types of learning situations. *J. educ. Psychol.*, 1960, 51, 208–212.

Beam, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, 51, 543–551.

Becker, S. L., & Dallinger, C. A. The effect of instructional methods upon achievement and attitudes in communication skills. *Speech Monogr.*, 1960, 27, 70–76.

Becker, S. L., Murray, J. N., & Bechtoldt, H. P. *Teaching by the discussion method.* Iowa City: State Univer. of Iowa, 1958.

Bendig, A. W. Ability and personality characteristics of introductory psychology instructors rated competent and empathic by their students. *J. educ. Res.*, 1955, 48, 705–709.

Bendig, A. W., & Hountras, P. T. Anxiety, authoritarianism, and student attitude toward departmental control of college instruction. *J. educ. Psychol.*, 1959, 50, 1–8.

Bills, R. E. Investigation of student centered teaching. *J. educ. Res.*, 1952, 46, 313–319.

Birney, R., & McKeachie, W. The teaching of psychology: A survey of research since 1942. *Psychol. Bull.*, 1955, 52, 51–68.

Bloom, B. S. Thought processes in lectures and discussions. *J. gen. Educ.*, 1953, 7, 160–169.

Bovard, E. W., Jr. Group structure and perception. *J. abnorm. soc. Psychol.*, 1951, 46, 398–405. (a)

Bovard, E. W., Jr. The experimental production of interpersonal affect. *J. abnorm. soc. Psychol.*, 1951, 46, 521–528. (b)

Briggs, L. J. Intensive classes for superior students. *J. educ. Psychol.*, 1947, 38, 207–215.

Brown, A. E. The effectiveness of large classes at the college level: An experimental study involving the size variable and the size procedure variable. *Univer. Iowa Stud. Educ.*, 1932, 7, 3.

Bryan, G. L., & Rigney, J. W. An evaluation of a method for shipboard training in operations knowledge. Tech. Rept. 18, Electronics Personnel Res. Group, Univer. of Southern California, 1956.

Burke, H. R. An experimental study of teaching methods in a college freshman orientation course. Unpublished doctoral dissertation, Boston Univer., 1955. (*Dissertation Abstr.*, 1956, 16, 77–78)

Burkhardt, Sara M. A study in concept learning in differential calculus. Unpublished doctoral dissertation, Columbia Univer., 1956.

Burtt, H. E., Chassel, L. M., & Hatch, E. M. Efficiency of instruction in unselected and selected sections of elementary psychology. *J. educ. Res.*, 1923, 14, 154–161.

Calvin, A. D., Hoffman, F. K., & Harden, E. L. The effect of intelligence and social atmosphere on group problem solving behavior. *J. soc. Psychol.*, 1957, 45, 61–74.

Calvin, A. D., McGuigan, F. J., & Sullivan, M. W. A further investigation of the relationship between anxiety and classroom examination performance. *J. educ. Psychol.*, 1957, 48, 240–244.

Cantril, H., & Allport, G. W. *The psychology of radio.* New York: Harper, 1935.

Carpenter, C. R. The Penn State Pyramid Plan: Interdependent student work study groupings for increasing motivation for academic development. Paper read at 14th Nat. Conf. on Higher Education, Chicago, March, 1959.

Carpenter, C. R., & Greenhill, L. P. An investigation of closed-circuit television for teaching university courses. Instructional Television Res. Rept., Project No. 1. University Park: Pennsylvania State Univer., 1955.

Carpenter, C. R., & Greenhill, L. P. An investigation of closed-circuit television for teaching university courses. Instructional Television Res. Rept., Project No. 2. University Park: Pennsylvania State Univer., 1958.

Carrier, N. A. The relationship of certain personality measures to examination performance under stress. *J. educ. Psychol.*, 1957, 48, 510–520.

Casey, J. E., & Weaver, B. E. An evaluation of lecture method and small group method of teaching in terms of knowledge of content, teacher attitude, and social status. *J. Colo.-Wyo. Acad. Sci.*, 1956, (4), 54.

Cheydleur, F. D. Criteria of effective teaching in basic French courses. *Bull. Univer. Wisconsin*, Serial No. 2783, August, 1945.

Churchill, Ruth. Preliminary report on reading course study. Yellow Springs, Ohio: Antioch Coll., 1957. (Mimeographed)

Churchill, Ruth, & Baskin, S. Experiment on independent study. Yellow Springs, Ohio: Antioch Coll., 1958. (Mimeographed)

Corey, S. M. Learning from lectures vs. learning from readings. *J. educ. Psychol.*, 1934, 25, 459–470.

Corman, B. The effect of varying amounts and kinds of information as guidance in problem solving. *Psychol. Monogr.*, 1957, 71, No. 2 (Whole No. 431).

Coulson, J. E. An experimental teaching machine for research at S.D.C. Tech. Memo. 416. Santa Monica, Calif.: System Development Corp., 1959.

Coulson, J. E., & Silberman, H. F. Results of initial experiments in automated teaching. Rept. SP-73. Santa Monica, Calif.: System Development Corp., 1959.

Craig, R. C. Directed vs. independent discovery of established relations. *J. educ. Psychol.*, 1956, 47, 223–234.

Cronbach, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, 12, 671–684.

Crum, D. An investigation of the relationships between students' expressions of satisfaction with certain aspects of the college classroom situation and their achievement on final examinations. Unpublished honors thesis, Univer. of Michigan, 1958.

Cutler, R. L., McKeachie, W. J., & McNeil, E. B. Teaching psychology by telephone. *Amer. Psychologist*, 1958, 13, 551–552.

Davage, R. H. The Pyramid plan for the systematic involvement of university students in teaching-learning functions. University Park: Div. of Academic Res. and Services, Pennsylvania State Univer., 1958.

Davage, R. H. Recent data on the Pyramid project in psychology. University Park: Div. of Academic Res. and Services, Pennsylvania State Univer., 1959.

Dawson, M. D. Lectures versus problem-solving in teaching elementary soil sections. *Science Educ.*, 1956, 40, 395–404.

Deignan, F. J. A comparison of the effectiveness of two group discussion methods. Unpublished doctoral dissertation, Boston Univer., 1955. (*Dissertation Abstr.*, 1956, 16, 1110–1111)

Della-Piana, G. M. Searching orientation and concept learning. *J. educ. Psychol.*, 1957, 48, 245–253.

Deutsch, M. An experimental study of the effects of cooperation and competition upon group processes. *Human Relations*, 1949, 2, 199–232.

DiVesta, F. J. Instructor-centered and student-centered approaches in teaching a human relations course. *J. appl. Psychol.*, 1954, 38, 329–335.

Dreher, R. E., & Beatty, W. H. An experimental study of college instruction using broadcast television. *Instructional Television Res. Proj.* No. 1. San Francisco: San Francisco State Coll., 1958.

Ebbinghaus, H. *Memory: A contribution to experimental psychology*. Trans. by H. A. Ruger & C. E. Bussenius. New York: Teachers Coll., Columbia Univer., 1913.

Edmonson, J. B., & Mulder, F. J. Size of class as a factor in university instruction. *J. educ. Res.*, 1924, 9, 1–12.

Eglash, A. A group discussion method of teaching psychology. *J. educ. Psychol.*, 1954, 45, 257–267.

Elliott, P. N. Characteristics and relationships of various criteria of college and university teaching. *Purdue Univer. Stud. higher Educ.*, 1950, 70, 5–61.

Evans, J. L., Glaser, R., & Homme, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book*. Washington, D.C.: Dept. of Audio-Visual Instruction, National Education Association, 1960. Pp. 446–451.

Evans, R. I., Roney, H. B., & McAdams, W. J. An evaluation of the effectiveness of instruction and audience reaction to programming on an educational television station. *Amer. Psychologist*, 1954, 9, 361–362.

Faw, V. D. A psychotherapeutic method of teaching psychology. *Amer. Psychologist*, 1949, 4, 104–109.

Fitch, M. L., Drucker, A. J., & Norton, J. A., Jr. Frequent testing as a motivating factor in large lecture classes. *J. educ. Psychol.*, 1951, 42, 1–20.

Flexman, R. E., Matheny, W. G., & Brown, E. L. Evaluation of the School Link and special methods of instruction. Urbana: Inst. of Aviation, Univer. of Illinois, 1950.

Fritz, M. F. *Survey of TV utilization in Army training*. Instr. Film Res. Rept. U.S. Naval Training Device Center, Office of Naval Research, SDC 530-01-0, December 31, 1952.

Gaier, E. L. Selected personality variables and the learning process. *Psychol. Monogr.*, 1952, 66, No. 17 (Whole No. 349).

Galanter, E. H. (Ed.) *Automatic teaching: The state of the art.* New York: Wiley, 1959.

Gibb, J. R. The effects of group size and of threat reduction upon creativity in a problem solving situation. *Amer. Psychologist*, 1951, 6, 324. (Abstract)

Gibb, Lorraine M., & Gibb, J. R. The effects of the use of "participative action" groups in a course in general psychology. *Amer. Psychologist*, 1952, 7, 247. (Abstract)

Glaser, R., & Homme, L. E. Relationships between the programmed textbook and teaching machine. Paper read at Air Force Office of Scientific Res. and Univer. of Pennsylvania Conf. on the Automatic Teaching of Verbal and Symbolic Skills, December, 1958.

Goldstein, A. A controlled comparison of the project method with standard laboratory teaching in pharmacology. *J. med. Educ.*, 1956, 31, 365–375.

Greene, E. B. Relative effectiveness of lecture and individual reading as methods of college teaching. *Genet. Psychol. Monogr.*, 1928, 4, 457–563.

Greenhill, L. P. New directions for communication research. *AV commun. Rev.*, 1959, 7, 245–253.

Gruber, H. E., & Weitman, M. Cognitive processes in higher education: Curiosity and critical thinking. Paper read at Western Psychological Ass., San Jose, Calif., April, 1960.

Guetzkow, H., Kelly, E. L., & McKeachie, W. J. An experimental comparison of recitation, discussion, and tutorial methods in college teaching. *J. educ. Psychol.*, 1954, 45, 193–209.

Haigh, G. V., & Schmidt, W. The learning of subject matter in teacher-centered and group-centered classes. *J. educ. Psychol.*, 1956, 47, 295–301.

Haines, D. B. Cooperative vs. competitive discussion methods in teaching introductory psychology. Unpublished doctoral dissertation, Univer. of Michigan, 1959.

Haythorn, W., Couch, A., Haefner, D., Langham, P., & Carter, L. The effects of varying combinations of authoritarian and equalitarian leaders and followers. *J. abnorm. soc. Psychol.*, 1956, 53, 210–219.

Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.

Heyns, R. W. Conference leadership which stimulates teamwork. *Michigan Bus. Rev.*, 1952, 4, 16–23.

Hirschman, C. S. An investigation of the small groups discussion classroom method on criteria of understanding, pleasantness, and self-confidence induced. Unpublished master's thesis, Univer. of Pittsburgh, 1952.

Hoban, C. F., Jr., & Van Ormer, E. B. *Instructional film research 1918–1950.* (Pennsylvania State Univer. Instructional Film Research Program) Port Washington, N.Y.: U.S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-19, 1950.

Hoehn, A. J., & Saltz, E. Effect of teacher-student interviews on classroom achievement. *J. educ. Psychol.*, 1956, 47, 424–435.

Hoffman, L. R. Homogeneity of member personality and its effect on group problem solving. *J. abnorm. soc. Psychol.*, 1959, 58, 27–32.

Horwitz, M. The veridicality of liking and disliking. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 191–209.

Hovland, C. I., (Ed.) *The order of presentation in persuasion.* New Haven, Conn.: Yale Univer. Press, 1957.

Hovland, C. I., & Mandell, W. An experimental comparison of conclusion-drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, 47, 581–588.

Hovland, C. I., Janis, I. L., & Kelley, H. H. *Communication and persuasion.* New Haven, Conn.: Yale Univer. Press, 1953.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. *Experiments in mass communication.* Princeton, N.J.: Princeton Univer. Press, 1949.

Hudelson, E. *Class size at the college level.* Minneapolis: Univer. of Minnesota Press, 1928.

Husband, R. W. A statistical comparison of the efficacy of large lecture vs. smaller recitation sections upon achievement in general psychology. *J. Psychol.*, 1951, 31, 297–300.

Jenkins, R. L. The relative effectiveness of two methods of teaching written and spoken English. Unpublished doctoral dissertation, Michigan State Univer., 1952. (*Dissertation Abstr.*, 1952, 12, 268)

Jensen, B. T. A comparison of student achievement under conditions of class attendance and non-attendance. *Coll. & Univer.*, 1951, 26, 399–404.

Johnson, D. M., & Smith, H. C. Democratic leadership in the college classroom. *Psychol. Monogr.*, 1953, 67, No. 11 (Whole No. 361).

Jones, H. E. Experimental studies of college teaching. *Arch. Psychol., N.Y.*, 1923, 10, No. 68, 5–70.

Kasten, D. F., & Seibert, W. F. *A study of televised military science instruction.* Television Project Rept. 9. West Lafayette, Ind.: Purdue Univer., 1959.

Katona, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

Katz, D. *Gestalt psychology.* New York: Ronald Press, 1950.

Keatinge, M. W. *Comenius.* New York and London: McGraw-Hill, 1931.

Kersh, B. Y. The adequacy of "meaning" as an explanation for the superiority of learning by independent discovery. *J. educ. Psychol.*, 1958, 49, 282–292.

Kishler, J. P. The effects of prestige and identification factors on attitude restructuring and learning from sound films. Port Washington, N.Y.: U.S. Naval Training Device Center, Office of Naval Research. Tech. Rept. SDC 269-7-10, 1950.

Klapper, Hope L. *Closed circuit television as a medium of instruction at New York University.* New York: New York Univer., 1958.

Knower, F. H. Experimental studies of changes of attitude. I. A study of the effect of oral argument on changes of attitude. *J. soc. Psychol.*, 1935, 6, 315–347.

Knower, F. H. Experimental studies of changes of attitude. II. A study of the effect of printed argument on changes of attitude. *J. abnorm. soc. Psychol.*, 1936, 30, 522–532.

Koenig, Kathryn, & McKeachie, W. J. Personality and independent study. *J. educ. Psychol.*, 1959, 50, 132–134.

Kopstein, F. F., & Roshal, S. M. Learning foreign vocabulary from pictures vs. words. *Amer. Psychologist*, 1954, 9, 407–408.

Krauskopf, C. J. The use of written responses in the stimulated recall method. Unpublished doctoral dissertation, Ohio State Univer., 1960. (*Dissertation Abstr.*, 1960, 21, 1953)

Kruglak, H. Experimental outcomes of laboratory instruction in elementary college physics. *Amer. J. Physics*, 1952, 20, 136–141.

Krumboltz, J. D., & Farquhar, W. W. The effect of three teaching methods on achievement and motivational outcomes in a how-to-study course. *Psychol. Monogr.*, 1957, 71, No. 14. (Whole No. 443)

Lahti, A. M. The inductive-deductive method and the physical science laboratory. *J. exp. Educ.*, 1956, 24, 149–163.

Landsman, T. An experimental study of a student-centered learning method. Unpublished doctoral dissertation, Syracuse Univer., 1950.

Lepore, A. R., & Wilson, J. D. An experimental study of college instruction using broadcast television. *Instr. Television Res. Proj.* No. 2. San Francisco: San Francisco State Coll., 1958.

Lewin, K. Group decision and social change. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (2nd ed.) New York: Holt, 1952. Pp. 330–344.

Lewin, K., Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created social climates. *J. soc. Psychol.*, 1939, 10, 271–299.

Lifson, N., Rempel, P., & Johnson, J. A. A comparison between lecture and conference methods of teaching physiology. *J. med. Educ.*, 1956, 31, 376–382.

Little, J. K. Results on use of machines for testing and for drill upon learning in educational psychology. *J. exp. Educ.*, 1934, 3, 45–49.

Longstaff, H. P. Analysis of some factors conditioning learning in general psychology. *J. appl. Psychol.*, 1932, 16, 131–166.

Lovell, G. D., & Haner, C. F. Forced-choice applied to college faculty rating. *Educ. psychol. Measmt*, 1955, 15, 291–304.

Lumsdaine, A. A., & Glaser, R. (Eds.) *Teaching machines and programmed learning: A source book.* Washington, D.C.: Dept. of Audio-Visual Instruction, National Education Association, 1960.

Lyle, E. An exploration in the teaching of critical thinking in general psychology. *J. educ. Res.*, 1958, 52, 129–133.

Macomber, F. G., & Siegel, L. *Experimental study in instructional procedures. Progress Rept. No. 1.* Oxford, Ohio: Miami Univer., 1956.

Macomber, F. G., & Siegel, L. A study of large group teaching procedures. *Educ. Res.*, 1957, 38, 220–229. (a)

Macomber, F. G., & Siegel, L. *Experimental study in instructional procedures. Progress Rept. No.* 2. Oxford, Ohio: Miami Univer., 1957. (b)

Macomber, F. G., & Siegel, L. *Final report on the experimental study in instructional procedures.* Oxford, Ohio: Miami Univer., 1960.

Maier, N. R. F., & Solem, A. R. The contribution of a discussion leader to the quality of group thinking. *Human Relations*, 1952, 5, 277–288.

Maize, R. C. Two methods of teaching English composition to retarded college freshmen. *J. educ. Psychol.*, 1954, 45, 22–28.

Maloney, R. M. Group learning through group discussion: A group discussion implementation analysis. *J. soc. Psychol.*, 1956, 43, 3–9.

Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 166–173.

Mann, F. C., & Hoffmann, L. R. *Automation and the worker.* New York: Holt, 1960.

Martin, J. R. Two-way closed-circuit educational television. Res. Rept. No. 941-1. Cleveland: Case Institute of Technology, 1957. (a)

Martin, J. R. Two-way closed-circuit educational television. Res. Rept. No. 941-1. Cleveland: Case Institute of Technology, 1957. (b)

Martin, J. R., Adams, R. B., & Baron, M. R. Studies in educational closed-circuit television. Res. Rept. No. 948-5. Cleveland: Case Institute of Technology, 1958.

May, M. A., & Lumsdaine, A. A. *Learning from films.* New Haven, Conn.: Yale Univer. Press, 1958.

McCollough, Celeste, & Van Atta, E. L. Experimental evaluation of teaching programs utilizing a block of independent work. Paper read at Amer. Psychol. Ass., Washington, September, 1958.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952. Pp. 46 ff.

McKeachie, W. J. Anxiety in the college classroom. *J. educ. Res.*, 1951, 45, 153–160.

McKeachie, W. J. Individual conformity to attitudes of classroom groups. *J. abnorm. soc. Psychol.*, 1954, 49, 282–289.

McKeachie, W. J. How do students learn? In R. M. Cooper (Ed.), *The two ends of the log.* Minneapolis: Univer. of Minnesota Press, 1958. Pp. 26–35. (a)

McKeachie, W. J. Motivating students' interest. In R. M. Cooper (Ed.), *The two ends of the log.* Minneapolis: Univer. of Minnesota Press, 1958. Pp. 36–39. (b)

McKeachie, W. J. Students, groups, and teaching methods. *Amer. Psychologist*, 1958, 13, 580–584. (c)

McKeachie, W. J. Appraising teaching effectiveness. In W. J. McKeachie (Ed.), *The appraisal of teaching in large universities.* Ann Arbor: Univer. of Michigan, 1959. Pp. 32–36.

McKeachie, W. J. *Teaching tips.* (4th ed.) Ann Arbor: Wahr, 1960.

McKeachie, W. J., & Hiler, W. The problem oriented approach to teaching psychology. *J. educ. Psychol.*, 1954, 45, 224–232.

McKeachie, W. J., & Solomon, D. Student ratings of instructors: A validity study. *J. educ. Res.*, 1958, 51, 379–382.

McKeachie, W. J., Lin, Y. G., Forrin, B., & Teevan, R. Individualized teaching in elementary psychology. *J. educ. Psychol.*, 1960, 51, 285–291.

McKeachie, W. J., Pollie, D., & Speisman, J. Relieving anxiety in classroom examinations. *J. abnorm. soc. Psychol.*, 1955, 50, 93–98.

McTavish, C. L. *Effect of repetitive film showings on learning.* (Pennsylvania State Univer. Instructional Film Research Program) Port Washington, N.Y.: U.S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-12, November, 1949.

Melton, A. W., Feldman, N. G., & Mason, C. N. *Experimental studies of the education of children in a museum school.* Washington, D.C.: Amer. Ass. Museums, Publ. No. 15, 1936.

Mertens, Marjorie S. *The effects of mental hygiene films on self regarding attitudes.* (Pennsylvania State Univer. Instructional Film Research Program) Port Washington, N.Y.: U.S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-22, July, 1951.

Miller, N. E. Scientific principles for maximum learning from motion pictures. *AV commun. Rev.*, 1957, 5, 61–113.

Moore, Mary R., & Popham, W. J. The role of extra-class student interviews in promoting student achievement. Paper read at joint session of Amer. Ass. for the Advancement of Science and Amer. Educ. Res. Ass., Chicago, December, 1959.

Mueller, A. D. Class size as a factor in normal school instruction. *Education*, 1924, 45, 203–227.

Nachman, M., & Opochinsky, S. The effects of different teaching methods: A methodological study. *J. educ. Psychol.*, 1958, 49, 245–249.

Novak, J. D. An experimental comparison of a conventional and a project centered method of teaching a college general botany course. *J. exp. Educ.*, 1958, 26, 217–230.

Pace, C. R., & Stern, G. G. *A criterion study of college environment.* Syracuse, N.Y.: Psychol. Res. Center, Syracuse Univer., 1958.

Parsons, T. S. A comparison of instruction by kinescope, correspondence study, and customary classroom procedures. *J. educ. Psychol.*, 1957, 48, 27–40.

Parsons, T. S., Ketcham, W. A., & Beach, L. R. Effects of varying degrees of student interaction and student-teacher contact in college courses. Paper read at Amer. Sociol. Soc., Seattle, Washington, August, 1958.

Patton, J. A. A study of the effects of student acceptance of responsibility and motivation on course behavior. Unpublished doctoral dissertation, Univer. of Michigan, 1955.

Paul, J. B. The length of class periods. *Educ. Record*, 1932, 13, 68–75.

Paul, J., & Ogilvie, J. C. Mass media and retention. *Explorations*, 1955, 4, 120–123.

Potter, S. *One-upmanship*. New York: Holt, 1955.

Pressey, S. L. A simple apparatus which gives tests and scores—and teaches. *Sch. & Soc.*, 1926, 23, 373–376.

Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *J. Psychol.*, 1950, 29, 417–447.

Rasmussen, G. R. Evaluation of a student-centered and instructor-centered method of conducting a graduate course in education. *J. educ. Psychol.*, 1956, 47, 449–461.

Remmers, H. H., & Brandenberg, G. C. Experimental data on the Purdue Rating Scale for Instructors. *Educ. Admin. Superv.*, 1927, 13, 519–527.

Remmers, H. H. Learning, effort, and attitudes as affected by three methods of instruction in elementary psychology. *Purdue Univer. Stud. higher Educ.*, 1933, 21.

Riley, J. W., Ryan, B. F., & Lifshitz, Marcia. *The student looks at his teacher*. New Brunswick, N.J.: Rutgers Univer. Press, 1950.

Rohrer, J. H. Large and small sections in college classes. *J. higher Educ.*, 1957, 28, 275–279.

Ross, I. C. Role specialization in supervision. Unpublished doctoral dissertation, Columbia Univer., 1956. (*Dissertation Abstr.*, 1956, 17, 2701–2702)

Ruja, H. Outcomes of lecture and discussion procedures in three college courses. *J. exp. Educ.*, 1954, 22, 385–394.

Runkel, P. J. Cognitive similarity in facilitating communication. *Sociometry*, 1956, 19, 178–191.

Russell, H. E. Inter-relations of some indices of instructor effectiveness: An exploratory study. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1951.

Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, 47, 810–817.

Scheidemann, Norma V. An experiment in teaching psychology. *J. appl. Psychol.*, 1929, 13, 188–191.

Schellenberg, J. A. Group size as a factor in success of academic discussion groups. *J. educ. Soc.*, 1959, 33, 73–79.

Seashore, C. E. Elementary psychology: An outline of a course by the project method. *Aims and Progress of Res.*, No. 153. Iowa City: Univer. of Iowa, 1928.

Seashore, S. E. Group cohesiveness in the industrial group. Ann Arbor: Survey Res. Center, Univer. of Michigan, Publ. No. 14, 1954.

Seibert, W. F. *A brief report and evaluation of closed-circuit television instruction in the first semester calculus course.* West Lafayette, Ind.: Audio-Visual Center, Purdue Univer., 1957.

Seibert, W. F. *An evaluation of televised instruction in college English composition.* Television Project Rept. 5. West Lafayette, Ind.: Purdue Univer., 1958. (Mimeographed)

Seibert, W. F., & Honig, J. M. *A brief study of televised laboratory instruction.* Television Project Rept. 8. West Lafayette, Ind.: Purdue Univer., 1959. (Mimeographed)

Siegel, L., Adams, J. F., & Macomber, F. G. Retention of subject matter as a function of large group instructional procedures. *J. educ. Psychol.*, 1960, 51, 9–13.

Skinner, B. F. Teaching machines. *Science*. 1958, 128, 969–977.

Slomowitz, M. A comparison of personality changes and content achievement gains occurring in two modes of instruction. Unpublished doctoral dissertation, New York Univer., 1955. (*Dissertation Abstr.*, 1955, 15, 1790)

Smith, D. E. P. Applicational transfer and inhibition. *J. educ. Psychol.*, 1954, 45, 169–174.

Smith, D. E. P., Wood, R. L., Downer, J. W., & Raygor, A. L. Reading improvement as a function of student personality and teaching method. *J. educ. Psychol.*, 1956, 47, 47–58.

Smith, H. C. Team work in the college class. *J. educ. Psychol.*, 1955, 46, 274–286.

Smith, W. F., & Rockett, F. C. Test performance as a function of anxiety, instructor, and instructions. *J. educ. Res.*, 1958, 52, 138–141.

Spence, R. B. Lecture and class discussion in teaching educational psychology. *J. educ. Psychol.*, 1928, 19, 454–462.

Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, 42, 183–188.

Stephan, F. F., & Mishler, E. G. The distribution of participation in small groups: An experimental approximation. *Amer. sociol. Rev.*, 1952, 17, 598–608.

Stephens, A. L. Certain special factors involved in the law of effect. *Abstr. doctoral Dissertations, Ohio State Univer.*, 1953, 64.

Stephens College. Courses to be taught over closed-circuit TV. *Stephens Coll. News Reporter*, 1955, 14, No. 4.

Stern, G. G. Congruence and dissonance in the ecology of college students. *Student Med.*, 1960, 8, 304–339.

Stern, G. G. Environments for learning. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 690–730.

Stern, G. G., Stein, M. I., & Bloom, B. S. *Methods in personality assessment*. Glencoe, Ill.: Free Press, 1956.

Stolurow, L. M. Teaching machines and special education. *Educ. psychol. Measmt*, 1960, 20, 429–448.

Stone, G. R. The training function of examinations: Retest performance as a function of the amount and kind of critique information. *USAF Pers. Train. Res. Cent. Res. Rept.*, 1955, No. AFPTRC-TN-55-8.

Stuit, D. B., Harshbarger, H. C., Becker, S. L., Bechtoldt, H. P., & Hall, A. E. *An experiment in teaching*. Iowa City: Dept. of Speech, State Univer. of Iowa, 1956.

Stuit, D. B., & Wilson, J. T. The effect of an increasingly well-defined criterion on the prediction of success at Naval Training School (Tactical Radar). *J. appl. Psychol.*, 1946, 30, 614–623.

Sturgis, H. W. The relationship of the teacher's knowledge of the student's background to the effectiveness of teaching. A study of the extent to which the effectiveness of teaching is related to the teacher's knowledge of the student's personal background. Unpublished doctoral dissertation, New York Univer., 1958. (*Dissertation Abstr.*, 1959, 19, 11)

Taylor, S. Helen. An experiment in classification of students in mathematics. *Math. Teacher*, 1931, 24, 414–423.

Tharp, J. B. Sectioning in Romance language classes at the University of Illinois. In *Studies in modern language teaching*. New York: Macmillan, 1930. Pp. 367–432.

Thistlethwaite, D. L. College environments and the development of talent. *Science*, 1959, 130, 71–76.

Thistlethwaite, D. L. College press and changes in study plans of talented students. Evanston, Ill.: National Merit Scholarship Corp., 1960.

Throop, J. F., Assini, L. T., & Boguslavsky, G. W. *The effectiveness of laboratory instruction in strength of materials by closed-circuit television*. Troy, N.Y.: Rensselaer Polytechnic Inst., 1958.

Timmel, G. B. A study of the relationship between method of teaching a college course in mental hygiene and change in student adjustment status. Unpublished doctoral dissertation, Cornell Univer., 1954. (*Dissertation Abstr.*, 1955, 15, 90)

Ullrich, O. A. An experimental study in ability grouping. Unpublished doctoral dissertation, Univer. of Texas, 1926.

Underwood, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, 64, 49–60.

VanderMeer, A. W. *Relative effectiveness of instruction by films exclusively, films plus study guides, and standard lecture methods.*

2 OF 5

ED 002948

1.0
4.5
5.0
5.6
6.3
7.1
8.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963

(Pennsylvania State Univer. Instructional Film Research Program) Port Washington, N.Y.: U.S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-13, July, 1950.

VanderMeer, A. W. *Effect of film-viewing practice on learning from instructional films.* (Pennsylvania State Univer. Instructional Film Research Program) Port Washington, N.Y.: U.S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-20, November, 1951.

Vinacke, E. W. Some variables in buzz sessions. *J. soc. Psychol.*, 1957, 45, 25–33.

Wakely, J. H., Marr, J. N., Plath, D. W., & Wilkins, D. M. Lecturing and test performance in introductory psychology. Paper read at Michigan Academy of Science, Ann Arbor, March, 1960.

Ward, J. Group-study vs. lecture-demonstration method in physical science instruction for general education college students. *J. exp. Educ.*, 1956, 24, 197–210.

Warren, R. A comparison of two plans of study in engineering physics. Unpublished doctoral dissertation, Purdue Univer., 1954. (*Dissertation Abstr.*, 1954, 14, 1648–1649)

Waterhouse, I. K., & Child, I. L. Frustration and the quality of performance. *J. Personnel*, 1953, 21, 298–311.

Watson, R. P. The relationship between selected personality variables, satisfaction, and academic achievement in defined classroom atmospheres. Unpublished doctoral dissertation, Univer. of Michigan, 1956.

Weitman, M., & Gruber, H. E. Experiments in self-instructed study: Effects on immediate achievement, performance of achievement and educational values. Paper read at Western Psychol. Ass., San Jose, Calif., April, 1960.

White, J. R. Methods in engineering laboratory instruction. *J. engng Educ.*, 1945, 36, 50–54.

Wieder, G. S. Group procedures modifying attitudes of prejudice in the college classroom. *J. educ. Psychol.*, 1954, 45, 332–344.

Wilke, W. H. An experimental comparison of the speech, the radio, and the printed page as propaganda devices. *Arch. Psychol., N.Y.*, 1934, 25, No. 169.

Williams, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.*, 1955, 51, 573–580.

Wispé, L. G. Evaluating section teaching methods in the introductory course. *J. educ. Res.*, 1951, 45, 161–186.

Wolfle, D. The relative efficiency of constant and varied stimulation during learning. *J. comp. Psychol.*, 1935, 19, 5–27.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

Zeleny, L. D. Experimental appraisal of a group learning plan. *J. educ. Res.*, 1940, 34, 37–42.

2

Motivation, Teaching Methods, and College Learning

W. J. McKEACHIE
University of Michigan

The quest for the golden fleece was no more arduous than the quest for measures predicting academic achievement. Almost from their inception intelligence and achievement tests proved to have significant correlations with criteria of college success, but these correlations seldom exceeded .5. Perhaps because of this quick success, our failure to find further predictive measures has been particularly disheartening. Almost everyone has been agreed that individual differences in motivation are responsible for much of the remaining variance in college grades. Considering the effort devoted to finding predictors of college success, it is astonishing that we still have no measure of motivation which has proved generally useful for college admissions officers or counsellors.

Previous research on the relative effectiveness of differing college teaching methods has also produced scanty results. Hundreds of experiments have been carried out comparing such methods as lecture vs. discussion, large vs. small classes, and television vs. live instruction, and have produced worthwhile evidence but few conclusive principles (McKeachie, in press).

One possible partial explanation for the meager findings in both these areas is that teaching methods affect different students differently. Students who profit from one method may do poorly in another, while other students may do poorly in the first method and well in the second. When we average them together

---

I am grateful to Drs. Patricia O'Connor and Charles Smith for helpful suggestions at various stages of the preparation of this manuscript.

we find little over-all difference between methods and no over-all effect of a single motivational measure.

The practical problems of predicting academic achievement and identifying effective teaching methods are not alone in pointing to the importance of studying interactions between student characteristics and teaching methods. There are also good theoretical reasons for looking at such interactions.

Cronbach (1957) has cogently argued for the values of combining experimental and correlational approaches in studying behavior. His arguments for the study of interactions of general psychological and personality variables are just as forceful when extended to the more molar level of interactions between personality and social psychological variables. In fact, Cronbach himself draws examples from education of the potential value of a combined experimental and correlational approach.

In addition to the arguments from the levels of practical necessity and theoretical exigency, an intermediate level of argument may be adduced. I would argue in extension of Cronbach's thesis that even though education research needs both more basic research and more applied research, its greatest present need is for research which tests out theoretical concepts and relationships in a naturalistic setting. If we attempt to apply any current behavior theory to education, we find that the gap between theory and educational practice is so great that educators can draw no clear implications from behavior theory. This is not the fault of the behavior theorist. He is quite right, I think, in studying whatever interests him, working in relatively simple, pure situations. (See Spence, 1959.) Even in these situations, the task of developing comprehensive theories and establishing quantitative laws is difficult enough to absorb all of the energy and talent available. To ask the behavior theorist to work out the applications of his theories is simply to draw him away from his central task. Nevertheless we do need researchers who with understanding of behavior theory undertake research in complex natural situations. Such research almost inevitably will need to find ways of categorizing or dimensionalizing situations, will need to establish which individual difference dimensions are most useful in determining the parameters of the learning theorists' curves, and will need to develop a taxonomy of types of learning.

I believe that research in this middle area between basic theory and practical problems may help both the theoretician and the practitioner. Such research must almost inevitably deal with interactions between situational and individual characteristics, and thus I deliberately chose not to call this contribution "Motivation and College Learning" or "Teaching Methods and College Learning" but rather "Motivation, Teaching Methods, and College Learning."

When one enters this middle ground of naturalistic research, it is immediately apparent that the formulae of the behavior theorists must be greatly expanded, and that in a single study one cannot examine all the possible variables affecting behavior. What dimensions are likely to be most useful to investigate? What should our independent variables be?

As a social psychologist, my thought first turned to social psychological variables as a starting point. Unfortunately I know of no social psychological theory which specifies the basic dimensions of groups and provides measures of them. Cattell (1948) has discussed a dimension he calls "group syntality" which includes a number of measures of group effectiveness and of characteristics of group members; group dynamics researchers have talked about dimensions like group cohesiveness (Cartwright & Zander, 1960); other theoreticians have suggested other variables. But none of these theories says much about how individuals differ in their behavior in a group which possesses a given amount of syntality, cohesiveness, or other characteristics. In most of these systems group members can be represented as empty circles indistinguishable from one another except for the roles individuals may fill for the group.

Turning from social psychological variables to personality variables, I encountered much the same difficulties. Despite heroic efforts to determine basic dimensions of personality, personality researchers seem in little agreement about what these dimensions are, and in many cases even after a trait or dimension has been identified there is little sure knowledge about what types of situations elicit this trait.

After considering a number of such alternatives, I settled on the approach taken by McClelland, Atkinson, Heyns, Veroff, and others (Atkinson, 1958) in developing theory and measures which take account not only of individual differences in achievement, affiliation, and power motivation but also of the situational cues which elicit these motives.

But motives and situational cues are not the only variables needed in predictions of college behavior. To place them in perspective, let me briefly sketch my conception of the major sources of variance in college achievement.[1]

I would begin with the student. Students enter college with a number of important, relatively stable motives learned at home, in their previous educational experiences, and through other experiences. There are not only individual differences in these motives but also differences in the relative strength of different motives within individuals. Students also possess skills and abilities useful in achieving goals in college. Some of these, such as verbal ability, are measured by the intelligence tests given for admission; others like reading ability, note-taking skill, study habits, ability to gain the instructor's sympathy, and ability to make friends, are not measured in our entrance battery. Unfortunately for my sort of research, there is a motivational factor of major importance common to all classes and students—grades. Whatever his motives for attending college, a student is motivated for grades. At the very least he must maintain an average high enough to remain in college, and in most cases his access to vocational or other goals depends to some extent upon achievement of higher grades. This is probably particularly true at Michigan where we have a student body highly selected on the basis of high school grades, with relatively few students anticipating inheritance of positions resulting from the family fortune, and with a social climate emphasizing academic achievement. Because most students are highly motivated for grades, the differential effects of differing teaching methods in eliciting such motives as need for achievement, need for affiliation, and need for power are difficult to detect. Nevertheless, if everyone were motivated equally we might expect the correlation between ability measures and grades to be much higher than it is. Even the best combinations of measures of ability and criterion measures leave a good deal of variance unaccounted for—enough at least to leave hope that motivational measures may contribute significantly.

---

[1] Dr. John W. Atkinson and Yi-Guang Lin have given me useful suggestions for this conceptualization but should not be held responsible for the use to which their ideas have been put.

The total college environment has much to do with activating these motives and skills. As Pace and Stern (1958) have shown, colleges differ markedly in their students' perceptions of what are important goals and appropriate instrumental behaviors to attain these goals. I suspect that Pace and Stern could also show from their data that there is an interaction at this level; i.e., students' own motives and skills influence their perceptions of the important cues in the college environment. Thus for Student A, the college environment may seem to present compelling cues for expectations of satisfaction from academic achievement and the use of skills in studying, while Student B in the same college perceives cues to expectation of satisfaction from extracurricular leadership achieved by using skills in making friends and fulfillment of social obligations.

In addition to such general expectations the student develops specific expectations about the likely satisfactions and appropriate instrumental behaviors in his college classes. The cues provided by his instructors thus may activate motives or skills which have previously been latent, may maintain the activation of motives or skills already active, may frustrate expectations previously elicited by the total environment, or may arouse uncertainty and anxiety about the likelihood of success or failure in reaching his goals.

The discussion this far probably implies my solution to the problem of identifying important social psychological or situational dimensions. This, of course, was also the problem which concerned Rotter in the 1955 Nebraska Symposium (Rotter, 1955). My solution to the problem was an indirect one. The steps were simply to choose apparently important human motives and then to categorize situations in terms of the cues they provide for activating these motives. This seems to me to be virtually the same technique as that used by Rotter except that my categorization is based upon the measures and research of the McClelland-Atkinson groups. I would, however, go beyond Rotter in suggesting that situations might also be usefully categorized in terms of their relevance for eliciting individual differences other than in motivation. For example, we might also categorize situations in terms of their relevance for individual differences in abilities, habits, or defense mechanisms.

At this point we have the student in a class in which cues to motive satisfaction and to pathways to the goal are present. What happens next?

Here our mode of discourse is misleading, for I must present in sequence several processes which are in fact proceeding simultaneously. At the same time that the individual's motives are being activated, he is choosing a strategy,[2] such as maximizing satisfaction or minimizing regret. Since the student is involved in several courses as well as in extracurricular, social, and other activities, his strategy must involve assessment of the potential rewards and punishments and their probabilities of attainment in all of these situations. Whatever his strategy, its choice also involves an assessment of his habits and skills in relation both to the possible satisfactions in the situation and to the types of behavior necessary to achieve these satisfactions. From these dual assessments comes an estimate of the amount and kind of satisfaction and the probability of attaining it in each class. This combination of subjective expectancies of affect and of probability of affect is what I call motivation.[3]

The individual's habits and skills thus enter into motivation in determining subjective probability. But they also enter into predictions of achievement again in their own right, for if we are to predict a grade we must assess the instrumental habits actually used by the student to achieve his goals as well as his real skill in their use.

---

[2] The words "choice" and "strategy" imply a much greater degree of consciousness and rationality than I wish. Clearly much of the process is unconscious and intuitive.

[3] I use "motive" to refer to the stable tendency to receive satisfaction from attaining a class of goals, and "motivation" to refer to the strength of the tendency to strive for a particular goal at a particular time.

Let me illustrate this by an example. Let us suppose that we have a student who has a strong achievement motive in a classroom in which the teacher provides cues to standards of achievement and incentives for achievement. This arouses expectancies of achievement satisfaction for the student, but the strength of his motivation will also be influenced by his estimate of the chances of achieving the standards set by the instructor. Atkinson (1958) has suggested that motivation for achievement will be highest when subjective probability is moderate, i.e., when the uncertainty of the outcome is at its maximum.[4] Two students with strong achievement motives may select the same instrumental habits to achieve their goals; e.g., they may both do their homework, but if they differ in subjective probability of success, their motivation will differ and we may expect differences in attentiveness and persistence and ultimately in achievement. Similarly two students may both have strong achievement motives and both have the same expectancies of success, yet they may differ in the instrumental habits they choose to attain success or in the skill with which they actually perform the instrumental responses chosen, and these differences may be reflected in their ultimate achievement.

The individual's habits and skills probably also enter our equation in a third term. In discussing the first variable, the college environment, I suggested that the total environment develops motivational and instrumental expectancies which the student brings into the classroom. It is easy to see that frustration is likely to be created when the student finds that his expected satisfactions are not likely to be gained in a particular class. The announcement of an especially rigorous grading policy is almost invariably greeted with groans of dismay and an upsurge in frustration-instigated behavior. But the instructor who blocks or de-emphasizes such usual instrumental responses as note-taking, memorizing a given textbook, and answering specific factual test items is likely to encounter even stronger resistance. It may simply be that the instructor who emphasizes different instrumental responses produces frustration because he is lowering student's subjective probability of success; it may be that frustration results because one's well-learned habits come to have secondary motivational values in themselves; or it may be that any marked change in expectancy is unpleasant. In any case, habit-skill expectancies enter into frustration, and frustration in turn may have certain responses associated with it which interfere with responses instrumental to learning.

---

[4] Although Atkinson has made his case only for achievement motivation, it is not improbable that a similar interaction might hold for other motives. In fact if it does not hold, it is pitiful to think of all the coyness that has been wasted by the folklore that man's ardor is enhanced by uncertainty.

At this point we have the student doing homework, taking lecture notes, making friends with the coed in the next seat, or performing other instrumental responses. If we were to stop right here, we would have research problems of sufficient complexity to keep us busy for years to come. (And now that I've made this analysis, I wish we'd at least done some preliminary research at this level.) But we have chosen as our dependent variables educational achievements, such as those summarized in course grades, and this means that there are several remaining possible slips between the cup and the lip.

For one thing, you will note that I have not yet considered learning in our equation. We have the student studying, listening to lectures, or participating in discussions, but how much these activities contribute to his learning will presumably depend upon such things as the distribution of trials, the pattern of reinforcements, the timing and amount of feedback, the relationship of the elements presented to one another, and the size of gaps between one step and the next. These add up to pointing to the importance of the skill of the teacher in programming the course to achieve his objectives. Once again, moreover, we must consider the probability that an effective learning program for some students may not be equally effective for others.

It is also important to note that since we are trying to assess the learning in an entire course, students' choice of instrumental behaviors and the motives which are activated are likely to change as a result of feedback received during the course. The teacher's behavior is important not only in the initial cues it provides to motives and habits, but also in the continuous shaping of instrumental responses and motives of students which goes on in addition to the type of learning assessed in course examinations.

By this point I hope we have considered the major classes of variables affecting the college student's learning, but we still have to assess that learning, and this introduces another set of variables. Obviously I should point to the technical problems of test construction, but I won't. Only briefly will I mention that teaching effective for one level of learning may not be effective in achieving other desirable outcomes (see, for example, McKeachie, 1959). But even more troublesome is the fact that all of the types of variables affecting the student's behavior in learning during the course also affect his behavior in the evaluation situation. Typically we use the course grade or the final examination as our principal criterion measure. Entering into the student's performance on the examination is not only his knowledge or other learning ostensibly assessed by the examination but also his motivation in the testing situation and his skills in taking examinations. One of the really oppressive thoughts which worries me is that some variable may work in opposite ways in the learning and examination situations. For example, students with a given level of anxiety about achievement might learn well, but in the examination situation, the cues to failure might be so great as to cause these students' performance to be disrupted. If we are really interested in learning, we would miss the importance of this variable. In addition, if course grade is our criterion measure, we have the additional problem of the differences in criteria used by instructors in awarding grades. The permissive instructor, for example, may reward class participation. Teachers differ in the range of behaviors they may observe as bases for grades. These are the sorts of problems which lead me to look for multiple criteria measures, but I am then faced with the additional problem of classifying and relating these measures.

One of the annoying things about naturalistic research is that the researcher typically lists so many variables that the reader or auditor leaves the field in confusion. I had thought that I might draw a diagram with arrows indicating interactions, but the arrows became so thick that I ran for cover and will simply try to list the categories of variables to which I have referred and to indicate the sorts of combinations I consider particularly important.

TABLE 1

Classes of Variables Interacting in Determining Achievement in a College Course

Environment	Immediate Situation	Student	Teacher	Achievement Test
College	Classroom cues	Motives, Habits, Skills	Patterning of reinforcements. Amount of feedback. Timing of feedback. Organization of content, etc.	1. Technical qualities. 2. Problem of multiple criteria with low or negative correlations. 3. Student motives and skills.

College and classroom cues × Motives × Habits = Motivation (and Frustration).
Motivation × Habits = Learning behaviors (e.g., studying, attending class).
Motivation × Learning behaviors × Teacher programming = Learning.
Learning × Motivation × Test-taking skills = Measured Achievement.

The discouraging thing about such an analysis is that one obviously can't study all of the possible interactions. What does one do? Well, he just jumps in at the points he thinks are most likely to be critical and does his best with what he can handle. As I've already indicated, I jumped into student motives and classroom cues, and while I've obtained measures in some of the other areas, I haven't tried to include them in my major analyses. Basically then, my work is an investigation of college student achievement as affected by the interaction between classes of cues found in college classrooms and the motives popularized by McClelland, Atkinson, and associates.

Now let me be more specific about how I try to work on these problems in an academic setting.[5]

---

[5] This research has been supported by grants from the Fund for Advancement of Education, Lilly Endowment, and U. S. Office of Education.

HYPOTHESES

Because our hypotheses about the affiliation motive are somewhat less complicated than our hypotheses about achievement motivation, I shall describe our procedures and results relevant to the affiliation motive before discussing our work on the power and achievement motives.

Basically our hypotheses have been of this form:

> The academic achievement of students with a strong affiliation motive will be relatively higher in classes high in affiliative cues than in classes with few affiliative cues, while the achievement of students with a weak affiliation motive will be relatively lower in classes high in affiliative cues than in classes with few affiliative cues.

Similar hypotheses may be stated for the criteria of student satisfaction and attitude change.

Let me review our general thinking using the affiliation motive as an example. We began by assuming that the affiliation motive is an important dimension of individual differences in college students. Classes differ in the extent to which they provide cues for this motive. Some instructors are warm, friendly, personally interested in each student; other instructors are subject- or self-oriented and elicit few expectations of affiliative satisfactions.

At this point, however, there is a gap in our theory of which I only gradually became aware. Our hypothesis assumes that students motivated to seek affiliative satisfaction in a class will achieve more than they would if no affiliative motivation were elicited. This fails to take account of the possibility that the teacher may not be the primary source of affiliative satisfaction. In fact one might expect affiliative satisfactions from other students to be more important than those from the instructor.

The second gap is that even when a student wants to win affiliative satisfaction from the teacher we are not sure that he will perceive academic achievement as being instrumental to such satisfactions. Quite conceivably, he might perceive apple-polishing, asking questions, or other behaviors as instrumental responses with a greater probability of arriving at his goal of affiliative satisfaction. Presumably his choice of these alternatives depends not only upon his own skills, but also upon constraints in the situation which reduce the probability of goal attainment for some alternatives.

My rationale for differential preformance of students low in the affiliation motive is less clear than that for those high in motive strength. Since we are using grades as a criterion, the simplest rationale is that students low in the affiliation motive are actually unaffected by the cues but that their relative standing will be higher when the other students are unmotivated and lower when the other students are motivated.

A somewhat more farfetched rationale, however, is also possible. McClelland, Atkinson, and others (1953) have shown that subjects scoring low in need for achievement include many individuals who have a strong fear of failure. Extending this to the affiliation motive, we might hypothesize that students scoring low on the test of the affiliation motive may fear rejection and become unproductively anxious in a class where the instructor takes a personal interest in students.

PROCEDURES

Sample

To test our hypotheses we[4] needed a sample of students differing in strength of motive and a sample of instructors differing in the cues emitted. Since we were primarily concerned as a dependent variable with *relative* achievement within a class, it would have been possible to use a number of sections of different courses. There were, however, some advantages to having several teachers trying to achieve the same goals, since this permitted us to compare over-all effects of differing methods. Yet at the same time there were clear advantages for generalization if conclusions could be based upon more than one course. Consequently we chose three multisection courses representing quite different types of content and objectives. They were second-year French, freshman mathematics, and general psychology. Thirty-one instructors participated—all of the experienced teachers in these courses.

---

[4] "We" here refers to Katherine Koenig, Anthony Kallet, Steven and Rae Kaplan, David Cross, Roland Tharp, and Jane Katherine Collins, the students who worked with me on the project described in the following section.

Within these instructors' sections all students were invited to take a group of tests designed to measure our independent variables. In psychology, the time spent was credited as fulfilling the requirement that each student participate in a psychological experiment. As a result almost every student participated. In French and mathematics we had less power but succeeded in testing some 80% of the men and 90% of the women. Of the 825 students in the three courses, 609 gave us usable responses on most tests. In our sample, there were 208 students in French, 188 in mathematics, and 213 in psychology. The subjects included 254 men and 355 women. Class size was typically 20 to 30, although one mathematics section had an enrollment of 60.

MEASURES

Motives and Ability

Our primary measure of motives was the thematic apperception measure of achievement, affiliation, and power motivation described by Atkinson (1958). The pictures used were those used in the national survey of mental health by Veroff, Atkinson, Feld, and Gurin (1960). Separate pictures were used for men and women. Our measure of anxiety was a questionnaire developed and used successfully at Stanford by Alpert and Haber (1960). As a measure of ability we used the scores on the American Council on Education Psychological Examination (1949 Form) which students had taken upon entering the University. We also developed measures of student expectations upon entering the course, of perceived instrumentality, of social class, religious background, and of other variables. None of these latter measures was included in the analyses to be reported here.

Methods

Our other independent variables were teaching methods. The most elegant design would have been to have each teacher teaching several classes, each of which was taught by a method specified to represent certain points on the dimensions we were studying. Since most instructors were teaching only one section, this wasn't possible, and even if it were, my previous experience with such manipulations has not been completely satisfying. There are a number of design problems which can not be solved by asking teachers to modify their behavior. For example, there is the problem of whether teachers can teach different sections of the same course by different methods and use each method adequately. In some cases efforts to force differentiation may result in such self-consciousness and artificiality that neither method is adequately represented. Another problem is that of Hawthorne (or reverse Hawthorne) effects created when students know that they are being experimented on. In any case, we simply let our instructors teach as they ordinarily would and used observers to place them on our dimensions.

We had two observers in each class. One of the observers used a modification of Bales' system of recording interactions (Mann, 1959). This system records each act in one of several categories. We collapsed some categories which were rare or unimportant in classroom observation. Rather than recording each student's acts separately, we treated the classroom as a two-person group made up of a teacher and students. We thus had student and teacher responses classified into eight categories:

1. Positive reactions.
2. Tension release.
3. Suggestions.
4. Orientation-opinion.
5. Asking for orientation or opinion.
6. Asking for support.
7. Tension.
8. Negative reactions.

The second observer was asked to look for behaviors which we expected to be particularly relevant for the motives we were studying. This observer thus noted such things as whether the instructor called students by name, students remained to chat with the instructor after class, and the instructor summarized the day's material at the end of the period. In addition, he noted instances of tardiness and leaving class before the end of the period.

Classes were usually observed three times by each observer.

Our third and most important source of information about classroom process was the student himself. Each student completed a twelve-item questionnaire about characteristics of the classroom. This questionnaire contained the three items with the highest factor loadings on each of four factors emerging from a factor analysis of a longer questionnaire. The four factors had to do with achievement cues, friendliness, organization of the content of the course, and organization of testing and grading.

Typical items were as follows:

Achievement Cues

The instructor sets high standards of achievement.
Students compete to do well in this class.

Friendliness (Warmth)

The instructor takes a personal interest in students.
The instructor calls students by name.

Organization (Structure)

The instructor summarizes the day's work.

Testing and Feedback

The instructor announces tests in advance.
Students know when their comments are right and wrong.

Instructors' Perception and Motives

A third group of independent variables were involved in our investigation but are of only indirect relevance to the work I will report in this paper. These variables included instructor motives, values, and instrumental preferences. To assess these we interviewed each instructor for two hours, asking him about his career goals, his aims in teaching, and his judgment about the importance of the various teacher behaviors included in our questionnaire. In addition, each instructor was given a modification of Elizabeth French's Test of Insight which was scored for the achievement, affiliation, and power motives.

Outcome Measures

We had three types of outcome measures. The most obvious one was the course grade. A second measure was the student's satisfaction, which was assessed by a questionnaire which may be summarized by quoting the last question: "Rate the contribution of this course to your education." On the scale below this item, 1 was superior, 2, very good, and so on to 5, poor. The third class of outcome measures included attitude scales—Thurstone type measures of attitudes toward language, mathematics, mental illness, and psychology.

Methods of Analysis

Our results were analyzed using Lancaster's technique (1949, 1950, 1951) for partition of chi square as programmed for the I.B.M. 650 and I.B.M. 704 computors by Thomas O'Brien.[7] Some additional analyses were made using analysis of variance or t tests. In constructing the tables for chi square, distributions on variables were split as close to the median as possible for the total group included in the analysis, or, in the case of variables for which we suspected curvilinarity, into equal thirds.

Since there were few men in French and few women in Mathematics, analyses for men were usually made for the Psychology and Mathematics courses, using Course as one of the variables. Similar analyses for women were made for the Psychology and French courses.

In the analyses for the affiliation and achievement motives, sections were differentiated into high and low "warmth" or high and low "achievement" cues in terms of the mean of all students in the section on the questionnaire of student perception of classroom characteristics. For analyses relating to the power motive, we used the observations on the Bales-Mann system. The score was the ratio of student-volunteered responses to teacher questions.

RESULTS

Do the motives we studied make a difference in learning in differing classroom situations? Our results indicate that they do. But like most research results, they are not so clear-cut and consistent that no more work is necessary; in fact, we are trying to replicate them in another study now going on.[8] Nevertheless, the

[7] Hays (personal communication) suggests that this technique does not provide an accurate estimate of the statistical significance of interactions when lower order interactions produce significant effects. In the results presented I have cross-checked such results with other statistical tests.

[8] This is a study with Robert Isaacson and John Milholland under contract with the U. S. Office of Education. They and other members of our research staff have contributed materially to the thinking represented in this paper.

results do, I think, increase our understanding of the motives we measured and of classroom situations. Let us examine each motive in turn.

The Affiliation Motive

The affiliation motive as measured by the TAT is presumed to involve "concern with establishing, maintaining or restoring a positive affective relationship with another person." We had hypothesized that students differing in affiliation motive would be affected differentially by the "warmth" or friendliness of the instructor. This hypothesis was supported by a significant interaction between warmth and affiliation motive in determining grades for all students in Psychology and for male students in both Mathematics and Psychology. None of the results for our other criteria were statistically significant. (See Table 2.) It is interesting to note, however, that in Psychology, which was rated as the course providing the most affiliation cues, students high in need affiliation rated the course as being more valuable than students low in the affiliation motive. This difference was not found in the other courses.

Other interactions add intriguing complexities to this relationship. For example, we found that warmth also interacted with sex in determining grades. For women, warmth produced higher grades; men, on the other hand, did more poorly in high warmth sections.

It is, of course, difficult to know whether need-affiliation scores for men and women are comparable. On the pictures we used, the women's scores tend to be higher than those of men, and Coffman (1961) finds that women's superiority on tests of verbal aptitude lies at least partly in their superior knowledge of words describing characteristics of people, while items about "things" tend to be easier for men. Folklore also supports the notion that interpersonal relationships are more important to women than to men. Thus it makes good sense that women should be favorably affected by affiliative cues. In fact it may be that the somewhat weaker interactions between affiliation motive and affiliative cues for women than for men are simply the results of a general feminine sensitivity to affiliation cues which tends to wash out the effects of individual differences in affiliation motive in women. Offering support for this speculation is the fact that even the women low in affiliation motive tended to prefer the classes with high affiliation cues.

TABLE 2

The Interaction of Warmth and the Affiliation Motive with Grades

Psychology Students' Grades

Warmth	*n* Aff.	A & B	C & D	
Hi	Hi	30	18	48
	Lo	17	31	48
Lo	Hi	17	27	44
	Lo	24	20	44
		88	96	184

Source	χ^2	df	p
n Aff. × Warmth	0.0	1	
n Aff. × Grade	.8	1	
Warmth × Grade	.1	1	
n Aff. × Warmth × Grade	8.5	1	.01

Excerpted from a larger analysis including three other variables: Sex, Instrumentality, and Intelligence.

Male Mathematics and Psychology Students' Grades

Warmth	*n* Aff.	A & B	C & D	
Hi	Hi	25	21	46
	Lo	10	21	31
Lo	Hi	17	33	50
	Lo	16	17	33
		68	92	160

Source	χ^2	df	p
n Aff. × Warmth	.004	1	
n Aff. × Grade	.153	1	
Warmth × Grade	.550	1	
Warmth × *n* Aff. × Grade	5.2	1	.05

Excerpted from a larger analysis including the variables Intelligence, Instrumentality, and Course.

Recent studies of public school teachers by Cogan (1958) and by Ryans (1960) indicate that a dimension characterized by friendly, integrative affiliative nurturant behavior is important in public school teaching. Teachers high in this dimension produce more student self-initiated work than teachers low in the dimension. Thus a similar teaching variable also appears to be important at other educational levels.

Need for Power

Let us now turn to the second of the motives for which we scored our TAT protocols—need for power. Veroff (1957), who developed the scoring system for power motive, describes the power motive as that "disposition directing behavior toward satisfactions contingent upon the control of the means of influencing another person." He found that students who scored high in power motive enjoyed leadership and recognition and were rated by instructors as high in argumentativeness and in frequency of trying to convince others of their point of view in the classroom.

One of the scores we had derived from our Bales-Mann observations was the proportion of student participations in discussion which were volunteered, rather than elicited by questions from the teacher. It seemed likely that a class in which the teacher encouraged students to volunteer their ideas freely would be one which would allow the student with high power motive to attempt to convince others of his point of view, and thus one in which he would be motivated. We thus hypothesized (weakly) that the student high in power motivation would achieve well in classes characterized by a high proportion of student volunteering. I characterize our hypothesis as "weak" because we were not confident that a student motivated for power would necessarily see study and academic achievement as instrumental to satisfaction of his motive.

The results relevant to these hypotheses were, however, encouraging. Male students high in the power motive do get better grades in classes in which student volunteering is prevalent than in classes in which it is uncommon. The converse is true for males low in power motive. (See Table 3.)

TABLE 3

The Interaction of *n* Power and Opportunity for Assertion in Determining Grades

(Subjects all men)

Assertion	*n* Power	Grades A & B	C & D	Total
Hi	Hi	25	30	55
	Lo	17	39	56
Lo	Hi	12	27	39
	Lo	19	22	41
		73	118	191

Source	χ^2	df	p
Assertion × *n* Power	0.0	1	
Assertion × Grade	0.1	1	
n Power × Grade	0.1	1	
Assertion × *n* Power × Grade	3.9	1	.05

Taken from an analysis which included the variables of Course and ACE.

Interactions with intelligence are interesting. The opportunity to present one's own point of view is apparently most important for male students in the top half of our group in intelligence, for we find a significant interaction between student-assertive teaching and ACE in determining grades. Males low in intelligence do poorly in classes permitting a high degree of student assertiveness, while high ACE students do well in such classes.

As you have probably noted, I have said in each case "male students." These findings do not emerge for females. Perhaps power motivation is not emphasized as an important aspect of socialization for girls, or perhaps assertiveness is not a behavior instrumental to satisfaction of power motivation for women.

The interaction of student-assertive teaching with intelligence is consonant with similar findings in other studies of superior performance by able students in smaller, more permissively led groups (e.g., Ward, 1956; Calvin, Hoffman, & Harden, 1957), so that again we have some outside support for regarding this as an important dimension of teaching.

TABLE 4

The Interaction of Intelligence and Opportunity for Assertion in Determining Grades

(Subjects all men)

Assertion	ACE	Grades A & B	Grades C & D	Total
Hi	Hi	30	31	61
	Lo	12	38	50
Lo	Hi	11	23	34
	Lo	20	26	46
		73	118	191

Source	χ^2	df	P
Assertion × ACE	3.4	1	
Assertion × Grade	0.1	1	
ACE × Grade	1.95	1	
Assertion × ACE × Grade	7.4	1	.01

The Achievement Motive

The third motive which we studied is one familiar to Nebraska Symposium devotees through the papers of McClelland (1955) and Atkinson (1954), the achievement motive or *n* Achievement. The very label, *achievement* motive, suggests that this motive should predict success in college. Many psychologists, however, have investigated the relationship between the achievement motive and grades with meager results. Atkinson (1958, p. 605) accounts for this result by the suggestion that most behavior in everyday life is overdetermined, i.e., that a number of motives are engaged so that predictions on the basis of a measure of any one motive is inevitably weak.

Atkinson's reasoning and the research which supports it cannot be gainsaid. But despite our recognition of the low probability of positive results, we included a measure of the achievement motive in our research design. One reason for our hope was that some of the studies reporting low correlations between the test of achievement motive and grades have used product-moment correlations or other statistics appropriate for linear relationships. There is much evidence that need-achievement scores derived in the manner developed by McClelland and his associates are nonlinear in their relationships with many other variables (Brown, 1953; McClelland & Liberman, 1949; McClelland, Atkinson, Clark, and Lowell, 1953, p. 269). If the relationship between need-achievement scores and grades is curvilinear, linear statistics would be unlikely to show the significance of the relationship.

But the more basic reason for reopening the question of the relationship between achievement motivation and grades is found in the rationale for this entire research program. If instructors differ in the cues they provide for motives, with affiliative cues dominant in some classrooms, achievement cues in other classrooms, and other cues in still others, the over-all grade-point average reflects a multiplicity of motive-arousing situations. Not only is achievement in any one class a goal for several motives and hence overdetermined in Atkinson's sense, but also in the grade-point average we are summarizing goal-achievement in situations differing in motive-arousal. Thus a low relationship between *n* Ach. scores and over-all grade-point average does not necessarily mean that *n* Ach. may not be a useful predictive measure in individual classes.

Our hypotheses about the achievement motive and achievement were complex. (The results were even more complex.)

First of all, we expected sex differences. Most of the positive results from the TAT measurement of the achievement motive have derived from studies of men. Women's scores tend to be higher than those for men. In some respects, they seem to be continually achievement-oriented rather than requiring achievement cues to activate motivation (Veroff, Wilcox, & Atkinson, 1953). Measures of intelligence predict grade-point average better for women than for men (at least at Michigan). This suggests that women are more uniformly motivated for academic achievement than are men. We thus hypothesized an interaction between sex, achievement cues, and the achievement motive in determining grades. For men our hypothesis was similar to those for the affiliation and power motives, i.e., that with more achievement cues, men high in the achievement motive would achieve better grades. For women, on the other hand, we expected that the optimal conditions for students high in the achievement motive would be in classes relatively low in achievement cues. Assuming that the Yerkes-Dodson law holds and that women are already at an optimal motivational level, we would expect an increase in achievement cues to result in a decrement in achievement.

A second complicating factor is that the measure of the achievement motive obtained from thematic apperception apparently is the resultant of two motives, one directed toward success, the other toward avoiding failure. Atkinson and Litwin (1960) have suggested that thematic apperception measures the motive to approach success while the Mandler-Sarason Test Anxiety Questionnaire measures the motive to avoid failure. O'Connor's results (1960) suggest that both scores represent interactions of the two motives. A low need-achievement score may indicate either a low need to achieve success or a strong need to avoid failure. Similarly a low anxiety score may represent low fear of failure, low need for success, or both.

With such ambiguities in the measures, you may well wonder why we used them at all. The answer is partly theoretical and partly simple curiosity. The body of theory developed by McClelland, Atkinson, and their associates is not only impressive in itself, but has been supported by a substantial corpus of research. It seemed to us that our investigation might help clarify the meaning of need for achievement and fear of failure by indicating the effects of classroom cues upon students differing in these motives.

Adding fear of failure to the hypotheses we have already suggested, we predicted that the achievement of students high in fear of failure, as measured by the Achievement Anxiety Scale (Alpert & Haber, 1960), would be higher in highly structured classrooms than in those where organization and structure were lacking. Let me explain our reasoning.

Whatever one's theoretical interpretation of the many studies relating anxiety to behavior, there is much evidence that a high level of anxiety is detrimental to complex learning or performance (e.g., see Spence 1956; Sarason, 1960), and we would therefore expect students high in anxiety to perform poorly in college. But anxiety also affects performance on intelligence tests, so that

Mandler and Sarason (1952) actually found students high in anxiety achieving higher grades than would have been predicted on the basis of their entrance tests. Their explanation is that the stress of entrance examinations is greater than that of achievement tests.

Anxiety is heightened by uncertainty (Raphelson, 1957; Sinha, 1950); so we might expect the anxious person to have the most difficulty in classes where the instructor's organization is poor and the students have little awareness of where they are going and what will happen next. On the other hand, in a well-structured course where the path to achievement is clearly defined, we might expect anxiety to have positive effects.[9] Atkinson (1958, p. 329) has suggested that in a constrained situation which the anxious person cannot avoid, his performance will be similar to that of the individual with high positive motivation for achievement.

Analysis of our data revealed significant interactions of the sort predicted but interactions which are not easily interpretable. For example, in psychology, achievement cues and the achievement motive interact in determining grades. But it is the group who score in the mid-third of the distribution of *n* Ach. scores who do best in the sections presenting fewer achievement cues. This is not unreasonable if this is a group high in fear of failure as suggested by Brown (1953), McClelland and Liberman (1949), and McClelland, Atkinson, Clark, & Lowell (1953). One would theorize that as achievement cues are increased, such students perform less well because they adopt defensive avoidance behaviors such as daydreaming or leaving the examination early as described by Atkinson and Litwin.

Unfortunately this relationship does not hold for the other courses. We do get a significant interaction between achievement cues and *n* Ach. in determining grades, but it is the high *n* Ach. students who do best in low cue sections (see Table 5). In fact the group for whom the difference is greatest is the men in the top third of the *n* Ach. distribution whose mean grade is 2.1 in the high cue sections and 2.6 in the low cue sections.

---

[9] Isaacson suggests that increased structure may simply make the cues to anxiety more apparent and affect performance detrimentally.

You will recall that we expected sex differences. These emerged in a significant three-way interaction between achievement cues, sex, and their need-achievement in predicting grades in psychology. Here, as we expected, men high in *n* Ach. did poorly in sections low in achievement cues. Women high in *n* Ach. did relatively better in low cue than in high cue sections. These interactions did not prove to be significant in other courses, but there was a significant interaction between achievement cues and sex, with men generally doing better in low cue sections than in sections high in cues for achievement, while for women the differences were slight. Actually the interaction between Course, Sex, Achievement cues, and Grades is also statistically significant. In Psychology the men do better in the high cue sections.

Ratings of course value indicate that *both* men and women high in the achievement motive prefer the classes with a low level of achievement cues; and students scoring low in *n* Ach. rate the high cue sections as more valuable.

Thus Psychology seems to be a maverick in these analyses, although results in the psychology course make sense in terms of our original hypotheses. Even with the deviant psychology course included, however, we find that over all the student with a high need for achievement does best in the sections with fewer achievement cues.

The simplest explanation of this in terms of our Yerkes-Dodson type of theorizing would be to suggest that there are enough cues in even the low cue sections to elicit the motivation of the student high in *n* Ach., and an increase in cues carried such students beyond the optimal level of motivation. An even more parsimonious explanaton might be to suggest that the high *n* Ach. student was motivated in the low cue sections, and the effect of higher achievement cues was simply to get the low *n* Ach. students motivated so that the high *n* Ach. students faced rougher competition.

I must admit that while embracing the solid virtues of parsimony, I have sometimes cast a wistful eye on the charms of fancy. So permit me to propose a third possible interpretation.

I would begin by distinguishing between extrinsic and in-

TABLE 5
The Interaction of Achievement Cues and the Achievement Motive in Determining Grades
(Subjects both men and women)

Ach. Cues	n Ach.	Grades A & B	Grades C & D	Total
Hi	Hi	68	82	150
	Lo	70	83	153
Lo	Hi	82	56	138
	Lo	58	84	142
		278	305	583

Source	χ^2	df	P
Ach. cues × *n* Ach.	0.0	1	
Ach. cues × Grades	1.2	1	
n Ach. × Grades	4.4	1	.05
Ach. cues × *n* Ach. × Grades	5.3	1	.05

trinsic rewards for achievement, a theme first presented by Harlow in the first Nebraska Symposium and revived by Koch and others in later symposia. For the student with high intrinsic motivation, rewards for achievement are internalized so that when he is placed in a situation where the task involves some competitive standard of achievement, he is motivated to do well and receives satisfaction when his achievement approaches or reaches his level of aspiration. Other students, however, may be dependent upon extrinsic rewards for achievement. Grades, praise, or some other form of recognition may be necessary for them to be motivated.

What happens to these two types of students in a situation in which the extrinsic rewards are emphasized? I would suggest that both types of students continue to work, but that for the intrinsically motivated student some of the zest is gone. Why is it that we sometimes refuse pay for a good deed we have done? Because the pay would rob us of the good feeling we gain from enhancing our own self-image.

Something of this sort may have happened in our study. In the sections characterized as strong in achievement cues, I suspect that the extrinsic rewards were prominent and that the student high in intrinsic achievement motivation may have felt driven. President Victor Butterfield of Wesleyan University has suggested[10] that these conditions may, in fact, actually break down intrinsic motivation. He believes that lack of curiosity and internalized motivation for learning is often the result of years of teaching which has so emphasized extrinsic incentives that the student has few opportunities for finding intrinsic satisfactions.

Empirical support for this position is found in Atkinson and Reitman (1956). They found that in a multi-incentive condition, subjects with high *n* Ach. worked fewer arithmetic problems than in a situation where achievement was the only incentive. Smith (1961) found even more dramatic differences in the same direction. In both studies subjects low in *n* Ach. did better with multiple incentives.

One of the allurements of this explanation is that it can handle the troublesome divergencies of the psychology course. Although the psychology course did not differ from the other courses in achievement cues, it was taught in a more permissive manner as measured by our student assertion index. Thus in Psychology we might expect a greater opportunity for intrinsic motivation to find satisfaction even in the sections with many achievement cues.

Now let us turn to the complications introduced by our attempt to measure fear of failure by an anxiety measure. You will recall that our basic hypothesis was that students high in achievement anxiety would perform better in highly structured than in less structured classes.

At the simplest level our results do not support this hypothesis, although there is some support in student ratings of course value. In psychology, students rated the classes with a high degree of structure as more valuable than those low in structure, and there was a significant interaction between structure and anxiety with the more anxious students differentiating more clearly in their preference for sections with a high degree of structure.

---

[10] Personal communication.

Some rather puzzling higher order interactions emerge. For example, there is a significant triple interaction between structure, sex, and anxiety when grades are the dependent variable (see Table 6). High anxiety women do well in highly structured classes and poorly in low structure as predicted. For men, however, it seems to be the classes with moderate structure which affect men high and low in anxiety differentially, with high anxiety men doing badly and low anxiety men doing well in such classes. To add to the confusion, the men in the mid-third of the distribution of anxiety scores achieve higher grades in highly structured sections than in less structured sections. This interaction simply confuses me.

Our most intriguing results come from the interactions of achievement motivation and anxiety with the degree to which the instructor made the steps to a good grade clear by announcing tests in class and indicating when student comments were right or wrong. We called this the "feedback" dimension. This dimension showed no interactions with anxiety alone, but was involved in a significant four-way interaction with sex, achievement motive, and anxiety in determining both grades and ratings of the value of the psychology course. To illustrate its complexity, let me point out some of the extreme cells.

Students who are high in achievement motive and low in anxiety received higher grades in sections low on the feedback dimension than in high sections, and their ratings of the value of the course are in the same direction. Female students low in achievement motive and low in anxiety didn't like and don't do well in sections low in feedback. Males low in achievement motive and low in anxiety did very well in sections high in feedback but rated the value of the course low.

These results seem to fit my fanciful ideas about extrinsic vs. intrinsic rewards. The high feedback section seems likely to be one where extrinsic rewards and punishments are stressed. The performance of the high *n* Ach., low anxiety student fits my conception that a student with a positive internalized motive will do best in a situation where intrinsic rather than extrinsic rewards are maximized. The low-low students on the other hand may be characterized as possessing weak intrinsic motives and

TABLE 6

The Interaction of Teaching Structure, Student Anxiety, and Sex in Determining Grades

(Subjects both men and women)

Sex	Anxiety	Mean Grades Structure Hi	Mean Grades Structure Mid	Mean Grades Structure Lo
Male	Hi	2.2	1.3	2.3
	Mid	2.7	2.2	1.7
	Lo	2.7	2.8	2.5
Female	Hi	3.0	2.5	2.0
	Mid	2.3	2.5	2.8
	Lo	3.3	2.7	3.0

$N=108$
$F=2.8$
$df=4$ and 54
$p=.05$

The other variable included in this analysis of variance was Course (mathematics, psychology, French).

hence requiring more extrinsic pushes and pulls. Admittedly any interpretations of such complex interactions are pretty shaky, but it's fun trying.

CONCLUSIONS

What do these results add up to?

To the obviously biased researcher, they add up to support for the original idea that teaching methods interact with student motives in determining the outcomes of education. Certainly the amount of variance which may be accounted for by any one of our present variables is not great, but it is significant.

A related conviction which has been strengthened by these results is that we need to make much greater use of multivariate analyses in our research. I must admit that I have difficulty interpreting some of the higher order interactions emerging from [illegible] analyses; moreover, I'm not sure that I trust them. Nevertheless, I believe that those of us doing research in social and educational psychology, particularly in naturalistic settings, have been in much the same position as the biologist before the discovery of the microscope. When we use more powerful analytic tools, we may, like Van Leeuwenhoek, first be awed by the "wee beasties" which swim up out of the murky placidity of our simple "no-significant-difference" studies. But in the long run, I am sure that we will develop the theoretical tools necessary to comprehend what we find.

Next I would point to the support for using individual difference variables as a key to the identification of important social psychological or educational dimensions. Undoubtedly there are social psychological variables which have uniform effects upon all sorts of people, but our progress in extending a list of these has been frustratingly slow, and I suspect that one reason for this is that dimensions keyed to differences between roles and individuals may be both more numerous and more powerful in predicting behavior in many social situations.

Finally I am relatively pleased with the outcomes in relation to our original problem of identifying important human motives in education. Those we chose were certainly not picked out of a hat, but their usefulness in an educational situation was somewhat doubtful, particularly when we considered the relatively low reliability of the best measures that were available. My regard for McClelland, Atkinson, Heyns, Veroff, and their associates is enhanced by finding that their variables and measures really do work—not as well as I should like, but at least enough to reinforce my own need for achievement.

REFERENCES

Alpert, R., & Haber, R. N. Anxiety in academic achievement situations. *J. abnorm. soc. Psychol.*, 1960, 61, 207-215.

Atkinson, J. W. Explorations using imaginative thought to assess the strength of human motives. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1954*. Lincoln: Univer. of Nebraska Press, 1954. Pp. 56-112.

Atkinson, J. W. (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

Atkinson, J. W., & Litwin, G. H. An experimental study on need-achievement and examination anxiety. *J. abnorm. soc. Psychol.*, 1960, 60, 52-63.

Atkinson, J. W., & Reitman, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, 53, 361-366.

Brown, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, 48, 469-476.

Calvin, A. D., Hoffman, F. K., & Harden, E. L. The effect of intelligence and social atmosphere on group problem solving behavior. *J. soc. Psychol.*, 1957, **45**, 61-74.

Cartwright, D., & Zander, A. *Group dynamics.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.

Cattell, R. B. Concepts and methods in the measurement of group syntality. *Psychol. Rev.*, 1948, **55**, 48-63.

Coffman, W. E. Sex differences in responses to items in an aptitude test. *Educational Testing Service Research Memorandum* RM-61-4, 1961.

Cogan, M. L. The behavior of teachers and the productive behavior of their pupils: II, Trait analysis. *J. exper. Educ.*, 1958, **27**, 107-124.

Cronbach, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, **12**, 671-684.

Lancaster, H. O. The derivation and partition of χ^2 in certain discrete distributions. *Biometrika*, 1949, **36**, 117-129.

Lancaster, H. O. The exact partition of χ^2 and its application to the problem of the pooling of small expectations. *Biometrika*, 1950, **37**, 267-270.

Lancaster, H. O. Complex contingency tables treated by the partition of χ^2. *J. Royal Statistical Soc.*, Series B 1951, **13**, 242-249.

Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.

Mann, R. The relation between personality characteristics and individual performance in small groups. Unpublished doctoral dissertation, Univer. of Michigan, 1959.

McClelland, D. C. Some social consequences of achievement motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. of Nebraska Press, 1955. Pp. 41-62.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

McClelland, D. C., & Liberman, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236-251.

McKeachie, W. J. *The appraisal of teaching in large universities.* Ann Arbor: Univer. of Michigan Extension Service, 1959.

McKeachie, W. J. The work of the teacher. In N. Sanford (Ed.), *The American college.* New York: Wiley, in press.

O'Connor, Patricia. The representation of the motive to avoid failure in thematic apperception. Unpublished doctoral dissertation, Univer. of Michigan, 1960.

Pace, R. C., & Stern, G. C. *A criterion study of college environment.* Syracuse University, Psychological Research Center, 1958.

Raphelson, A. C. The relationship between imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13-18.

Rotter, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. of Nebraska Press, 1955. Pp. 245-269.

Ryans, D. G. *Characteristics of teachers.* Washington: American Council on Education, 1960.

Sarason, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403-415.

Sears, Pauline S. Problems in the investigation of achievement and self-esteem motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1959.* Lincoln: Univer. of Nebraska Press, 1959. Pp. 265-339.

Sinha, A. K. P. Experimental induction of anxiety by conditions of uncertainty. Unpublished doctoral dissertation, Univer. of Michigan, 1950.

Smith, Charles. Situational determinants of the expression of achievement in thematic apperception. Unpublished doctoral dissertation, Univer. of Michigan, 1961.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W. The relation of learning theory to the technology of education. *Harvard educ. Rev.,* 1959, 29, 84-95.

Veroff, J. Development and validation of a projective measure of power motivation. *J. abnorm. soc. Psychol.,* 1957, 54, 1-8.

Veroff, J., Atkinson, J. W., Feld, Sheila, & Gurin, G. The use of thematic apperception to assess motivation in a nation-wide interview study. *Psychol. Monogr.,* No. 499, 1960, 74, 12.

Veroff, J., Wilcox, Sue, and Atkinson, J. W. The achievement motive in high school and college age women. *J. abnorm. soc. Psychol.,* 1953, 48, 108-119.

Ward, J. Group-study vs. lecture-demonstration method in physical science instruction for general education college students. *J. exp. Educ.,* 1956, 24, 197-210.

Comments on Dr. McKeachie's Paper

Seymour B. Sarason

For several years Dr. Kenneth Davidson and I have been utilizing classroom observations in our studies of anxiety in children. I can therefore testify not only to the complexity of making, scoring, and analyzing such observations but to the fact that Dr. McKeachie has utilized an approach which begins to do justice to the complexity. Given a complex situation, it is too frequently the case that one handles the problem either by essentially denying its complexity or by fractionating it to the point where one's efforts have the taint of irrelevance.

Dr. McKeachie's paper gives me an excuse for airing a strong and long-lived opinion: I feel that it was a mistake of no small proportions when after World War II clinical psychology was allied to psychiatry. The first of the two major reasons for this conclusion is my opinion that, as a field, clinical psychology could have made a greater contribution to our society by operating within the school-community rather than the psychiatric setting

—an opinion which is truly irrelevant to the purposes of this symposium. The second reason for this opinion was the conviction that in the school setting one was faced with and could study almost every important problem in psychology. Motivation, learning, group processes, personality and culture, child development—these and other problem areas not only can be studied in the school setting (be it elementary school, high school, or college), but can be studied in a controlled way. I certainly do not want to minimize the difficulties of controlled studies in such a setting. They are, however, far from insuperable, and this leads me to conclude the school as a psychological laboratory has decided advantages over the clinic and hospital, and the conventional psychological laboratory. I think Dr. McKeachie's paper is a good example of what I mean, i.e., in his present paper, as well as in the future studies he will be doing, Dr. McKeachie studies concurrently variables of central importance to a variety of psychologists.

One of the points in the paper with which I would take issue concerns the assumption that students come to college motivated to get good grades. I seriously question the universality of such a motivation. I think there are a fair number of students who enter college uninterested in or unconcerned about grades, but motivated rather by the youthful phantasy that they are approaching (and will conquer) the fortress of knowledge and wisdom—the secrets of life and living will be theirs. I am sure there are other students who enter college consciously not motivated for grades but motivated to go to college because it is, so to speak, the thing to do. In any event, until empirical evidence indicates otherwise, I think it reasonable to assume that not all students are motivated for grades, at least when they enter college. If this turns out to be the case, it clearly must be taken into account in attempts to predict grades as well as in how one interprets the significance of grades as a measure of how much learning has taken place.

Dr. McKeachie's assumption that all students come to college motivated for grades obscures for him what I think is an unstudied but intriguing social psychological problem. Although I assume that not all students are motivated for grades when they enter college, I think the college culture influences or changes student motivation so that not too long after entrance practically all students are consciously motivated for grades. I would be surprised (and very much disappointed) if the college's emphasis on grades was not resented by many students. I would expect that there are at least two kinds of students: both are motivated for grades, but in one student such a motivation raises no conflict with his other values, while the other student resents having to feel motivated for grades. In other words, not only do students vary in motivation for grades when they enter college, but they vary in how this motivation interacts with and is influenced by the values of the culture. Dr. McKeachie has, I think, oversimplified the "motivation for grades variables" and bypassed the problem of the influence of a culture on motivation, values, and personality.

The above comments were in part prompted by some data I collected when we were studying test anxiety in Yale College students. In three successive years I administered our test anxiety scale to all of the introductry psychology classes. At the end of the instructions I said something like this: "When you have answered all the questions, please hand in your questionnaire to me. However, I would like to request, and this is completely optional, that on the back of the questionnaire you write your opinions of the role of tests or examinations in higher education. Write as little or as much as you want. Remember, you do not have to write anything." What happened? First, a small minority of each class wrote nothing. Second, of the majority who did answer the question, many wrote to the extent of several long paragraphs. Third, there was near unanimity that to interpret the results of tests (on which grades are based) as reflecting what a student has learned in the course—or how he can utilize what he learned—was unfair and absurd. (It should be pointed out that almost all who wrote their opinions were referring to objective tests.) I am sure that some of these replies were from students who were motivated for grades at the same time that they resented such motivation in themselves and derogated the culture which forced such a motivation upon them. To sum up: I think Dr. McKeachie, who comes to this problem as a social psychologist, is using a methodology which I could characterize as clinical and psychometric, the clinical reflected in his measures of motivation and the psychometric in his measure of college learning. I find myself in the awkward position of maintaining that it might be more fruitful if his methodology reflected more of a social psychological–cultural anthropological orientation.

Related to the above point is Dr. McKeachie's handling of sex differences; witness his discussion of the possible significance of the finding that the affiliation scores of women are higher than those of men.

> . . . Folklore also supports the notion that interpersonal relationships are more important to women than men. Thus it makes good sense that women should be favorably affected by affiliative cues. In fact it may be that the somewhat weaker interactions between affiliation motivation and affiliative cues for women than for men are simply the results of a general feminine sensitivity to affiliation cues which tends to wash out the effects of individual differences in affiliation motivation in women. Offering support for this speculation is the fact that even the women low in affiliation motivation tended to prefer the classes with high affiliation cues.

Why is it so common in psychological research that sex differences are not predicted, analysis then reveals them, and "folklore" is resorted to for an explanation?* Why is it that psychological theory (with the notable exception of psychoanalytic theory) has never come to grips with what folklore considers obvious? If the sexes do differ significantly in the structure of their motivational complex, it not only affects the kinds of problems Dr. McKeachie is working on—affects his problems in terms of choice of motivational variables and procedures for their measurement—but forces him, it seems to me, to give more attention to the ways in which these motivational differences are influenced by the academic culture. Unless one takes seriously the culture in which a problem is being studied, interpretation of one's findings may indeed be misleading. For example, is Dr.

---

*A truly embarrassing example of this can be found in our recent book in the chapter on sex differences (S. B. Sarason, K. S. Davidson, F. F. Lighthall, R. R. Waite, and B. K. Ruebush: *Anxiety in Elementary School Children: A Report of Research*).

McKeachie assuming that his findings have generality and that he would expect similar findings if his studies were conducted in an all-female or all-male college culture?

Comments on Dr. McKeachie's Paper

David Birch

Attempts to investigate the contribution of some of psychology's recognized variables in settings as complex as the university classroom are certainly to be commended. As a bystander to this kind of endeavor I'm sure my "subjective probability of success" is considerably farther from .5 than is Professor McKeachie's.

It seems to me, however, that the recognition that important effects may lie at the level of interactions rather than with simple effects is noteworthy. The classroom is clearly one of complex stimuli, heterogeneous subjects, and complicated criteria, all of which invite interaction effects if appropriate assessments can be made.

The variables of need achievement, need affiliation, and need power have for some time held considerable intuitive appeal as being important in the classroom setting. Somehow it seems that variables of this type ought to be useful tools for analysis. Whether or not these intuitive notions are adequately defined by existing methodologies is a further question.

Certainly the results found with the three need measures are encouraging. Interesting interactions seem to result. The future requirement is surely that of replication and extension, as Professor McKeachie has indicated.

THE MICHIGAN PROJECT:

RESEARCH ON CHARACTERISTICS OF EFFECTIVE COLLEGE TEACHING

Most previous educational research has failed to reveal differences in the effect of differing teaching methods. One of the reasons for this is that evaluation techniques have often been inadequate, but another reason is that teaching methods may have different effects upon different types of students. Very able students, for example, may profit from teaching methods which are less effective for students of less ability. Similarly, students with certain personality characteristics may benefit more from one method of teaching than from others. Since different students may react differently to any teaching method, lumping all students together could hide the actual effects of the method.

The research reported herein was intended to accomplish three tasks:

1. Devise instruments for the evaluation of student learning in a prominent undergraduate course, introductory psychology.

2. Use these instruments to measure:

   a. the effects of differing teacher characteristics and of various student characteristics upon learning.

   b. differential outcomes of the course as a function of the interactions among method, teacher personality, and student personality.

3. Apply the results and findings in the selection and training of new college teachers. This objective was not achieved. The project did not get carried to the point where this task could be undertaken, but considerable progress was made in description of teacher behavior, and some progress made in identifying characteristics of effective teachers.

II-A. MEASURING COGNITIVE ABILITIES

John E. Milholland

If it is true that there are differential effects of various student-teacher method combinations on overall student performance, it seems likely also that different facets of what is to be learned in a course might be subject to these interactions. This point of view has characterized the approach to outcome measures used in this project.

The Committee on Criteria for the First Course* of the Division on Teaching of the American Psychological Association had already been active in this area. In 1957-58 they sent inquiries about objectives for the first course in psychology to teachers in 300 colleges listed in the then current directory of the American Psychological Association as having three or more members on their staff. The questionnaire listed 16 possible objectives, asked the respondents to indicate which ones they tried to accomplish, to tell how they measured the accomplishment, and invited them to list any additional objectives they had. The objectives included in the questionnaire appear in Appendix A. The first nine were taken from the Taxonomy of Educational Objectives (Bloom, 1956), Number 10 from Dressel and Mayhew (1954) and the remaining six were suggested by the Committee.

One hundred thirteen replies were received. Practically no measurement techniques were provided, and a number of the respondents expressed interest in any such the Committee might work out. Some additional objectives were suggested, but after scrutiny by the Committee all but a very few of them could be subsumed under the 16 originally proposed.

At this point Professor Paul Dressel materially assisted the Committee by obtaining from the members of an informal association of university examiners copies of examinations used in their institutions for the first course in psychology. These, together with a large item file available at The University of Michigan provided an item pool of considerable magnitude. Items were circulated among members of the Committee for study and assignment to relevant objectives.

The contract between the U.S. Office of Education and The University of Michigan for research on the characteristics of effective college teaching went into effect on January 1, 1960. This contract made it possible to hold a four-day meeting of the Committee, June 8-11, 1960. The Committee was augmented by W. J. McKeachie and R. L. Isaacson, co-directors of the research project, and by Virginia Lickey and Norman Weissberg, research assistants for the project. The test items that had been evaluated by the Committee were reviewed and refined and some new items added.

---

*At this writing the Committee consists of Donald M. Johnson, David R. Krathwohl, Slater E. Newman, and John E. Milholland, chairman. Former members are Donald N. Elliott and Sumner C. Hayward.

At this meeting the group was able to locate or produce enough satisfactory items for try-out for only the first 9 of the 16 objectives listed in Appendix A. They thought, however, that No. X, Critical Thinking, and No. XI, Disposition and ability to seek the explanation for human behavior in more than superficial characteristics, would be to some extent covered by items classified under other categories. Eight test forms (1960 Forms A-H) of 50 items each were prepared.

Forms A and B of the test were given preliminary try-out with summer session classes at Michigan State University, North Carolina State College, and The University of Michigan. Mean scores (number right) for the various classes were between 28 and 36, with standard deviations between 4 and 7. Probable errors of measurement were about 2 points. These results were interpreted as encouraging, and preparations were made to use all eight forms of the tests more extensively the following year.

Each of the forms A-H was given as a pre-test to between 150 and 200 students in introductory psychology in the fall semester of 1960. At the end of the semester each student took a form different from the one he had taken as a pre-test, so that gains could be studied. Motivational conditions were far from ideal, since the students knew the tests were experimental. Nevertheless, gain scores were of the order of 3 or 4 points, and the post-test performances were used for item analysis.

As a result of the item analysis some items were modified, others dropped, and the number of forms reduced from 8 to 4 (labeled A-D). Items for Objective V, Analysis of Organizational Principles, failed to survive the item analysis, so that only 8 objectives remained. The distribution of items among them is shown in Table 1.

The tests were administered as pre-tests and as post-tests in the spring semester of 1961. About a fifth of the students did not finish the pre-tests, due in the most part to the fact that a number of instructors took so much time to explain the purpose and the nature of the tests that working time was cut too short. Pre-test data for only those who finished the test were used, which may have introduced some upward bias. There were thus 583, out of a total of 747, scores available. Results of the pre-tests and post-tests are shown in Table 2.

TABLE 1

DISTRIBUTION OF ITEMS AMONG OBJECTIVES IN FORMS A-D OF THE CRITERIA TEST

SPRING SEMESTER, 1961

Objective	Number of Items				
	Form A	Form B	Form C	Form D	Total
I. Interpretation	6	10	5	9	30
II. Application	10	7	8	8	33
III. Analysis of Elements	6	6	6	6	24
IV. Analysis of Relationships	6	5	9	6	26
VI. Production of a Plan	3	3	3	3	12
VII. Derivation of a Set of Abstract Relations	7	7	7	7	28
VIII. Judgments in Terms of Internal Evidence	1	1	0	0	2
IX. Judgments in Terms of External Criteria	11	11	12	11	45
Total	50	50	50	50	200

TABLE 2

PRE-TEST AND POST-TEST SCORES ON CRITERIA TEST FORMS A-D

SPRING SEMESTER, 1961

	Form							
	A		B		C		D	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
N	145*	187	151*	191	143*	177	144*	182
Mean	23.3	30.0	23.9	29.8	26.8	31.0	21.7	31.1
SD	5.62	5.47	6.11	6.13	5.94	6.40	6.37	5.74
Reliability**		.77		.73		.78		.79

*Includes only students who finished the test.

**Rulon Formula.

The gains in mean scores are about a standard deviation, but this may be attenuated by whatever upward bias in the pre-test scores was engendered by the restriction to students who finished. Disregarding whether or not the pre-test was finished, there were 719 students who took the test both times. Their mean pre-test score was 22.8, with a standard deviation of 6.58; their mean post-test score was 30.5, with a standard deviation of 5.98.

The data in Table 2 seemed to justify the conclusion that the four forms of the test were nearly enough equivalent statistically to be used as parallel forms and that they were measuring abilities that were being developed or improved in the course. Reliabilities were adequate for use with groups.

A principal axes factor analysis with varimax rotation was performed on the inter-item Phi-coefficients (fourfold point correlations) for each form of the test. Factor loadings were generally small and did not conform to the subscores that were supposed to represent the various objectives. It was decided, therefore, that only total scores on the Criteria Tests would be used in the analyses that were to be performed on the Spring, 1961, data.

The post-test had been administered at the time provided for the course final examination, but many teachers did not count the scores in their grades, and those who did gave them little weight. The post-test scores correlated .50 with course grades and .48 with scores on the American Council on Education Psychological Examination (ACE) which the students had taken when they entered the University as freshmen. The ACE correlated .35 with course grades, and the correlation of post-test scores and grades, with ACE partialed out, was .46.

Although the difficulties attending the administration of the pre-test suggested that using gains from pre-test to post-test would be of doubtful value as an outcome measure, some correlations were computed. The gains correlated .15 with course grades; and, as is frequently the case, there was a negative correlation (-.52) with pre-test scores. When pre-test scores were partialed out, the correlation of gains with grades rose to .39. The gain score adjusted for pre-test score did not appear to be clearly interpretable, however, and no subsequent analyses involving it were carried out.

The Criteria Test measure most widely used as an outcome measure was the mean post-test score for all the students of an instructor. This "instructor mean post-test score" was sometimes used as a raw score, sometimes converted to a standard score, and sometimes the measure was the residual error of estimate of prediction from ACE score. This last measure will be called "post-test adjusted for ACE."

A multiple-choice test of 15 items designed to sample the content of the course was also administered to 548 students with the course final examination. The correlation of this test with the Criteria post-test was

.36. Its correlation with the separate forms of the Criteria Test were A: .36, B: .31, C: .37, and D: .44. The correlation of course grades with the 15-item test was .26 for a sample of 193 men and .27 for a sample of 224 women. Corresponding correlations of the test with ACE were .27 and .24.

In the fall semester of 1961 the elementary psychology course at Michigan was split into two parts: Psychology 100, Psychology as a Natural Science; and Psychology 101, Psychology as a Social Science. A natural question was whether the Criteria Tests favored one course or the other, so the four forms were administered in both at the end of the semester. Results are shown in Table 3. Although mean scores were generally 2 or 3 points lower than those shown in Table 2, there seemed to be no significant bias against either course. Apparently whatever was lost by splitting into two courses was equally apportioned.

TABLE 3

CRITERIA TEST SCORES OF STUDENTS IN
PSYCHOLOGY 100 AND PSYCHOLOGY 101
END OF FALL SEMESTER, 1961

	Form and Course									
	A		B		C		D		All Forms	
	100	101	100	101	100	101	100	101	100	101
N	139	151	136	150	136	148	137	148	548	597
Mean	27.6	27.9	27.8	26.3	26.6	28.7	28.2	28.9	27.5	27.9
SD	6.73	5.75	6.27	6.23	6.87	6.42	7.00	5.37	6.75	6.05

Each student had taken two forms of the test and thus an opportunity was provided to correlate subscores for each objective across forms. The correlations are shown in Table 4. The distributions of interform correlations for the same objectives and intraform correlations for different objectives are compared in Table 5. There is some tendency for the correlations for the same objectives to be the higher, since 40% of them are above .30, as compared with 22% of the intraform coefficients. Likewise, 40% of the interform correlations are below .21, as compared with 49% of the intraform ones. Nevertheless, the relationships generally do not indicate that the subscores are measuring clearly differentiated and homogeneous abilities.

TABLE 4

CORRELATIONS* BETWEEN SUBSCORES FOR THE SAME OBJECTIVE ON DIFFERENT FORMS OF THE TEST: CRITERIA TESTS ADMINISTERED FALL SEMESTER, 1961**

Objectives	Forms Involved in the Correlation					
	AB	AC	AD	BC	BD	CD
I. Interpretation	18	35	42	24	27	17
II. Application	27	20	15	45	27	32
III. Analysis of Elements	45	38	35	38	39	34
IV. Analysis of Relationships	05	30	22	31	19	19
VI. Production of a Plan	-04	-07	03	03	08	20
VII. Derivation of a Set of Abstract Relations	38	39	26	49	10	35
VIII. Judgments in Terms of Internal Evidence	05	(Objective not included in Forms C and D)				
IX. Judgments in Terms of External Criteria	28	48	23	16	16	42
Total Score	63	70	63	62	55	70

*Decimal points omitted.
**N's varied from 115 to 267.

TABLE 5

DISTRIBUTION OF CORRELATIONS* AMONG SUBSCORES FOR DIFFERENT OBJECTIVES WITHIN THE SAME FORM AND FOR THE SAME OBJECTIVE IN DIFFERENT FORMS: CRITERIA TESTS ADMINISTERED FALL SEMESTER, 1961**

r	Form of Test					Correlations for the Same Objectives in Different Forms
	A	B	C	D	Total	
41-50			4		4	6
31-40	7	5	3	3	18	11
21-30	8	7	4	9	28	9
11-20	3	7	9	8	27	9
Below 11	10	9	1	1	21	8
Total	28	28	21	21	98	43

*Decimal points omitted.
**N's varied from 115 to 267.

It was thought that some of the difficulty might lie in the fact that some subtests consisted of a rather small number of items (see Table 1). It was decided, therefore, to reduce the number of forms to two, each having 60 items measuring six objectives with subtests of 10 items each.

A continuing problem in the development of the test had been the reluctance of the instructors in the course, most of whom were teaching fellows with little or no previous teaching experience, to have it contribute as a substantial factor in determining grades. In the spring semester of 1962, however, the three project directors were involved in conducting a section of 130 students in Psychology 101. This section was taught by a succession of instructors, each being responsible for the work of a week or two. Each student took two forms of the Criteria Test as his final examination, which was to count for 25% of his grade in the course. Scores on the Criteria Tests were found to correlate .43 with the average of scores on the other tests given in the course. The mean score was 34, which was about a standard deviation above that for the fall group. A number of reasons could be advanced for the superiority, but all of them would be speculative. At any rate, the students apparently took the test seriously, conditions of administration were uniform and satisfactory, and it was decided to use item analysis data from this group as a basis for choosing items for the two new forms, X and Y.

In allocating items to Form X or Form Y primary consideration was given to balancing content and item style. In addition efforts were made to have the mean item difficulty for each subtest be the same in the two forms. The distributions of item difficulties are shown in Table 6.

In the fall semester, 1962, each student in Psychology 100 and Psychology 101 took both forms of the test as a part of his final examination. The test was to affect grades in the course, but the extent of the effect was left up to the instructors, and there was considerable variability. Score statistics are given in Table 7. The performance in the two courses seemed to be similar enough to warrant combining the data, and this was done for computing the means and standard deviations of subtest scores, shown in Table 8. Statistically the two forms seem nearly enough alike to be considered equivalent.

Correlations among subtest scores are presented in Table 9, and the distributions of interform and intraform correlations in Table 10. The correlation between total scores for the two forms was .69, which may be taken as the reliability of either form of the test.

TABLE 6

DISTRIBUTION OF ITEM DIFFICULTIES FOR FORMS X AND Y OF CRITERIA TESTS.

DATA FROM SECTIONS 14 AND 26 OF PSYCHOLOGY 101

SPRING, 1962

Percent Passing	Objective and Test Form												Total	
	I		II		III		IV		VII		IX			
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
95-99		1							1				1	1
90-94		1								2				3
85-89	1							1	1	3	1		3	4
80-84	1	1				2	2		1	2			4	5
75-79	1	2		1	2			1	1				4	4
70-74	2		2		2	1	2	1	1				9	2
65-69		1	1	2	1	2		3	1	1	1	1	4	10
60-64	2		2	3	2	1	2		2		1	1	11	5
55-59			3	1		2	3	2	1		1	2	8	7
50-54	2			1	1			1			2	2	5	4
45-49	1	1		1							1	1	2	3
40-44		2			1	1	1				2	2	4	5
35-39		1	1					1	1				2	2
30-34			1		1	1						1	2	2
25-29				1						2	1		1	3
20-24														
Mean	66.5	65.5	57.5	58.6	61.4	61.2	64.9	64.9	69.4	72.7	54.2	51.2	62.4	62.4

TABLE 7

SCORES ON FORMS X AND Y OF CRITERIA TESTS

FALL SEMESTER, 1962

	Psychology 100		Psychology 101	
	Form X	Form Y	Form X	Form Y
N	364	364	1085	1085
Mean	37.6	36.6	36.9	36.1
SD	6.74	6.86	6.81	6.95

TABLE 8

SUBTEST SCORES ON THE CRITERIA TESTS

FALL SEMESTER, 1962

Objective	Mean		Standard Deviation	
	Form X	Form Y	Form X	Form Y
I. Interpretation	5.6	5.8	1.88	1.74
II. Application	6.1	5.4	1.68	1.81
III. Analysis of Elements	6.9	6.6	2.17	2.12
IV. Analysis of Relationships	6.8	5.9	1.70	1.62
VII. Derivation of Abstract Relations	6.7	6.9	1.87	1.97
IX. Judgment by External Criteria	4.9	5.5	1.80	1.79
Total	37.0	36.2	6.88	6.99

TABLE 9

CORRELATIONS AMONG SUBTEST SCORES OF CRITERIA TEST FORMS X AND Y

FALL SEMESTER, 1962

Objective	Form	Objectives											
		I		II		III		IV		VII		IX	
		X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
I. Intepretation	X		25	32	25	30	32	28	26	28	32	24	30
	Y	25		21	23	24	29	22	21	20	22	18	26
II. Application	X	32	21		27	34	32	31	18	23	25	19	29
	Y	25	23	27		26	31	27	21	20	30	21	30
III. Analysis of Elements	X	30	24	34	26		56	28	21	25	32	22	30
	Y	32	29	32	31	56		29	27	29	37	24	37
IV. Analysis of Relationships	X	28	22	31	27	28	29		24	27	26	20	30
	Y	26	21	18	21	21	27	24		23	22	15	23
VII. Derivation of Abstract Relations	X	28	20	23	20	25	29	27	23		40	18	22
	Y	32	22	25	30	32	37	26	22	40		17	32
IX. Judgment by External Criteria	X	24	18	19	21	22	24	20	15	18	17		25
	Y	30	26	29	30	30	37	30	23	22	32	25	

TABLE 10

CORRELATIONS BETWEEN SCORES ON CORRESPONDING SUBTESTS IN FORMS X AND Y OF THE CRITERIA TEST AND BETWEEN SCORES ON SUBTESTS FOR DIFFERENT OBJECTIVES

Correlation	Corresponding Subtest	Different Subtests	
	Different Forms	Same Form	Different Forms
.55-.59	1		
.50-.55			
.45-.49			
.40-.44	1		
.35-.39		2	
.30-.34		8	7
.25-.29	3	8	9
.20-.24	1	10	10
.15-.19		2	4
Total	6	30	30

Only two objectives: III. Analysis of Elements; and VII. Derivation of a Set of Abstract Relations have subtests that seem to be measuring something unique. It may be that the logical structure of the Taxonomy of Educational Objectives does not conform to the psychological pattern of responses to the test items. We have thus far been unable to locate any trait validity data from research elsewhere, except for content analyses, on tests constructed to measure objectives listed in the Taxonomy.

A factor analysis of the subtest intercorrelations was performed, with the results shown in Table 11. The large first factor in the unrotated matrix, accounting for more than four times as much variance as any other factor, indicates a substantial general factor in the test performance. This may be why we have not been successful in developing distinct subtests.

The rotation of axes does not provide much help for interpretation. Objectives III and VII show up as Factors C and E, respectively, but beyond this there is little structure. The typical situation is that when one member of a cross-forms pair of subtests loads strongly on a factor, its partner does not. This kind of a finding casts doubt on the validity of judgmental processes for constructing items to measure somewhat subtle cognitive objectives.

In the spring semester of 1963 Form X of the test was given as a part of the final examination in Psychology 100 and 101. The results are shown in Table 12, and it can be seen that the means and standard deviations are very close to those shown in Table 7 for the fall semester.

TABLE 11

FACTOR LOADINGS (PRINCIPAL AXES WITH VARIMAX ROTATION) OF SUBTEST SCORES OF CRITERIA TEST FORMS X AND Y

FALL SEMESTER, 1962

Subtest (Objective)	Form	Unrotated Factors						h^2	Rotated Factors					
		1	2	3	4	5	6		A	B	C	D	E	F
I. Interpretation	X	61	-06	14	-10	06	-20	44	06	35	22	32	29	29
	Y	50	11	27	45	-25	32	70	72	-09	21	33	03	10
II. Application	X	57	19	-16	-09	47	-15	65	04	72	35	05	05	07
	Y	55	14	12	00	22	51	65	65	41	06	-10	21	08
III. Analysis of Elements	X	65	19	-50	13	-16	-21	79	06	16	85	07	12	10
	Y	70	13	-37	18	-22	-09	73	21	11	78	13	20	12
IV. Analysis of Relationships	X	57	-02	16	-07	47	-10	59	13	68	06	26	18	09
	Y	48	-28	41	44	-01	-40	83	11	11	11	88	12	04
VII. Derivation of Abstract Relations	X	54	-59	-02	-29	-14	01	74	-02	09	10	19	82	12
	Y	62	-41	-17	-18	-13	27	70	25	11	26	00	75	02
IX. Judgment by External Criteria	X	44	41	38	-50	-39	-18	95	08	06	10	04	07	96
	Y	61	18	09	00	03	19	45	44	32	25	07	16	25
	λ	3.95	.92	.89	.84	.82	.79	8.22	1.29	1.44	1.72	1.13	1.50	1.14

TABLE 12

STATISTICS OF CRITERIA TEST SCORES, FORM X

SPRING SEMESTER, 1963

Course	N	Mean	SD
Psychology 100	281	37.67	7.07
Psychology 101	550	37.38	6.94
Total	831	37.48	6.99

In Psychology 101 there had also been an objective and an essay test on the course content administered. Scores on these tests, the measures obtained from the TAT, Criteria Test scores, scores on the University's freshman test battery, and course grades for the 264 students for whom complete data were available were factor analyzed.* A principal-axes analysis with varimax rotation was used, with the results shown in Table 13. The two highest communalities, .91 each, for the Criteria Test and Essay Test totals are, of course, artifacts of the inclusion of the respective subtests in the analysis. The communalities and loadings for the subtests are likewise inflated by their presence in the total. It is possible, nevertheless, to see from Table 13 something of the nature of the factorial structure of the Criteria and Essay Tests.

Factor A picks up variance from the objective test, high school rank, SAT, the Criteria Test, and course grade, but not from the Essay Test. Factor B does contain an Essay Test loading, as well as some from the objective test, one subtest of the Criteria Test, and course grade. Taken together, Factors A and B account for a third of the variance in course grades and thus may represent achievement in general. The other loadings are consonant with this interpretation.

Factor D may be of some interest because of its loadings on the Reading Test and SAT-V. The .33 loading of the Criteria Test reassures us that the test is not predominantly one of verbal intelligence coupled with reading skill.

The remaining factors seem not to merit detailed interpretation, being practically doublets.

The Criteria Test emerges as a good test of achievement in psychology, so far as total score is concerned, but we cannot say that the subtest scores represent the objectives they were designed to measure, or in fact, with two possible exceptions, anything consistent and differentiable.

---

*The preparation and scoring of the objective and essay tests and the factor analysis were carried out by Karl Zinn as a part of his doctoral dissertation.

TABLE 13

FACTOR LOADINGS (PRINCIPAL AXES WITH VARIMAX ROTATION) OF
VARIOUS MEASURES FOR 264 STUDENTS IN PSYCHOLOGY 101
SPRING SEMESTER, 1963

	Factor											
	A	B	C	D	E	F	G	H	I	J	K	h²
Criteria Test, Form X												
I. Interpretation	31	11	07	26	-01	03	03	16	-02	-02	64	62
II. Application	33	40	-00	25	01	-06	-15	35	-20	-04	-06	57
III. Analysis of Elements	38	03	05	18	-06	00	05	-01	-05	01	-02	65
IV. Analysis of Relationships	31	25	-08	25	08	01	11	09	13	-06	16	52
VIII. Derivation of Abstract Relations	26	08	02	14	01	10	05	68	02	-03	12	59
IX. Judgment by External Criteria	11	10	-12	18	-26	15	02	10	63	08	-04	46
Total	58	27	-03	33	-01	06	03	37	10	-06	22	91
Objective Content Test	35	40	-06	24	-07	19	10	13	-14	07	07	49
Essay Test												
Application	00	86	-04	08	04	00	04	00	04	-02	03	76
Detecting Assumptions	11	26	-02	10	-07	63	01	13	05	-01	08	51
Producing a Plan	02	51	37	12	08	57	06	00	-02	01	-05	48
Analysis of Relations	02	30	08	11	01	01	77	02	00	02	08	70
Total	11	76	03	15	00	37	43	05	03	00	03	91
Thematic Apperception Test												
Need Achievement	07	-04	30	-13	14	15	16	02	-11	01	02	12
Need Affiliation	00	06	10	-06	54	03	04	02	-04	11	-07	33
Need Power	12	-07	06	01	-54	04	07	00	-02	-11	-05	33
Student's Percentile Rank in High School Graduating Class	58	18	-13	10	02	21	05	04	-07	15	00	31
College Board Scholastic Aptitude Test												
Verbal	37	16	-04	58	-04	09	10	08	-09	-05	08	60
Mathematical	67	08	-30	24	-01	15	02	-08	-13	-24	20	51
Cooperative Reading Test												
Speed	18	12	03	80	-02	07	06	08	09	-09	02	72
Accuracy	10	25	-04	79	00	04	03	04	-03	05	10	70
Opinion Attitude and Interest Survey												
Achiever Personality	07	13	-05	03	00	05	21	01	-02	41	07	51
Intellectual Quality	20	04	-02	44	-05	-01	08	11	-14	-43	16	49
Creative Personality	-06	-05	28	14	-14	-08	02	10	-05	-36	-04	28
Social Science Interest	03	00	63	07	02	01	07	04	-15	-21	07	51
Physical Science Interest	06	02	-68	10	-03	-08	02	04	-10	-04	-02	50
Biological Science Interest	-15	-05	-13	-10	19	-03	12	-01	01	55	-01	45
Course Grade	50	41	18	15	00	12	10	19	-11	08	16	48
λ	2.77	2.44	1.30	2.90	.69	1.07	.90	.88	.51	1.09	.65	14.80

References

Bloom, B. S. (ed). Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: Longmans, Green, 1956.

Dressel, P. L. and Mayhew, L. B. General Education: Explorations in Evaluation. Washington: American Council on Education, 1954.

APPENDIX A

OBJECTIVES OF THE FIRST COURSE IN PSYCHOLOGY

1. INTERPRETATION: When given a communication the student can identify and comprehend the major ideas which are included in it as well as understand their interrelationships. Illustrations: The ability to grasp the thought of work as a whole at any desired level of generality; the ability to distinguish among warranted, unwarranted, or contradicted conclusions drawn from a body of data; the ability to interpret various types of social data; ability to make proper qualifications when interpreting data.

2. APPLICATION: Given a problem new to the student, he will apply the appropriate abstraction without having to be prompted as to which abstraction is correct or without having to be shown how to use it in that situation. Illustrations: The ability to apply social science generalizations and conclusions to actual social science problems; apply principles of psychology in identifying the characteristics of a new social situation.

3. ANALYSIS OF ELEMENTS: Illustrations: The ability to recognize unstated assumptions; skill in distinguishing facts from hypotheses; the ability to distinguish factual from normative statements; skill in identifying motives and in discriminating between mechanisms of behavior with reference to individuals and groups; ability to distinguish a conclusion from statements which support it.

4. ANALYSIS OF RELATIONSHIPS: Illustration: Ability to recognize what particulars are relevant to the validation of a judgment; ability to recognize which facts or assumptions are essential to a main thesis or to the argument in support of that thesis; ability to check the consistency of hypotheses with given information and assumptions; ability to distinguish cause-and-effect relationships from other sequential relationships; ability to distinguish relevant from irrelevant statements in an argument; ability to detect logical fallacies in arguments.

5. ANALYSIS OF ORGANIZATIONAL PRINCIPLES: Illustration: Ability to see the techniques used in persuasive materials, such as advertising, propaganda, etc.

6. PRODUCTION OF A PLAN, or a proposed set of operations. Illustrations: Ability to propose ways of testing hypotheses; ability to integrate the results of an investigation into an effective plan or solution to solve a problem.

7. DERIVATION OF A SET OF ABSTRACT RELATIONS: Illustrations: Ability to formulate appropriate hypotheses; ability to formulate a theory of learning; ability to perceive various possible ways in which experience may be organized to form a conceptual structure.

8. JUDGMENTS IN TERMS OF INTERNAL EVIDENCE, such as logical accuracy, consistency, etc. Illustrations: Ability to assess general probability of accuracy in reporting facts from the care given to exactness of statement, documentation, proof, etc.; the ability to apply given criteria to the judgment of the work; the ability to indicate logical fallacies in arguments.

9. JUDGMENTS IN TERMS OF EXTERNAL CRITERIA: Illustrations: The comparison of major theories, generalizations, and facts about particular cultures; skills in recognizing and weighing values involved in alternative courses of action; the ability to distinguish between technical terminology which adds precision to a text by permitting more appropriate definition of terms, and that which merely replaces a common name by an esoteric one.

10. CRITICAL THINKING (problem-solving aspects): Ability to recognize the existence of a problem; ability to define the problem; ability to select information pertinent to the solution; ability to recognize assumptions bearing on the problem; ability to make relevant hypotheses; ability to draw conclusions validly from assumptions, hypotheses, and pertinent information; ability to judge the validity of the processes leading to the conclusion; ability to evaluate a conclusion in terms of its application.

11. The disposition and the ability to seek the explanation for human behavior in more than superficial characteristics. Illustrations: Recommending the amelioration of behavioral adjustment problems through remedying underlying, rather than precipitating, causes; recognition of the various kinds of behavior which may occur in response to basic human motives; understanding the symbolism of many behaviors.

12. The development of tolerance toward deviant individuals. Illustrations: Understanding and acceptance of research findings about racial and cultural differences; tolerance and concern for mental illness; appreciation of the problems of mental deficiency and mental superiority.

13. Appreciation and support of scientific and scholarly research. Illustrations: Appreciation and acceptance of scientific method as a problem-solving technique; recognition of the value of basic research; recognition of the value of the scholar in society.

14. Ability to make objective judgments about one's own behavior, to "see ourselves as others see us."

15. Personal adjustment: Ability to deal with one's own problems more effectively. Illustration: Ability of a person to recognize defense mechanisms when he himself uses them.

16. Interest in the content of psychology. Illustrations: Intention to take more courses in psychology.

II-B. ATTITUDINAL STUDIES

Dodd H. Bogart

It is a common notion in contemporary thought that part of the solution to the American racial problem lies in the educational process. The objective of tolerance is especially evident in the general area of psychology. The introductory textbook of psychology is rare indeed which does not discuss prejudice, the authoritarian personality, and racial problems. Nor is there customarily any doubt about the author's hope with respect to the direction of attitudinal change on the part of the students. Prejudice, it is clear, is to be reduced by education generally and the study of psychology in particular.

Nor is the reduction of racial prejudice the only objective of this nature which may be observed in the course in psychology. It is clear that tolerance and understanding for the mentally diseased, the physically handicapped, and the criminal can also be the intent of the instructor of psychology, the textbook, or the individuals taking the course. Certainly there are reasons to suppose that the study of psychology should influence student's attitudes whether this is intended or not. For psychology provides perspectives for viewing the determinants of human behavior which do not support beliefs of innate racial superiority, which do not lend themselves readily to an explanation of mental illness in terms of sin, which do not accommodate entirely punitive approaches to the problem of crime.

In view of the expectation that psychology does and should be instrumental in attitude change, it is interesting that attempts to measure this change have not been more successful (see Jacob, 1957). It is also interesting that much has yet to be done in the way of developing instruments for measuring belief systems related to racial problems, mental illness, crime and the role of psychology in human affairs in general. The Project on the Characteristics of Effective College Teaching took the assessment of some of these attitudes as a legitimate problem in the study of the classroom in introductory psychology. Certain traditional measures were employed and others developed. It is instructive to follow the selection and use of these measures in chronological order.

1. THE ADJECTIVE CHECK LIST

Sommer and Killian (1954) developed a list of adjectives which they found would discriminate prejudiced from unprejudiced subjects. They summarize their results as follows:

> The unprejudiced group prefers that the Negro act liberal, persistent, neighborly; in a general extroverted, self-assertive manner. The prejudiced group, on the other hand, prefers that the Negro act quiet, obedient, conservative, or, in a general introverted, ancillary manner (Sommer and Killian, 1954, p. 229).

This list (see forms in Appendix A) was administered during the summer session, 1960, to two classes of introductory psychology at both The University of Michigan and North Carolina State College. The two forms, i.e., the one asking for a rating of a person and the other for a rating of a Negro, were randomly assigned within each class. No independent measure of prejudice was obtained from these subjects. The analysis concerned whether the adjectives would be differentially responded to when the object to be judged was identified as "a Negro" and when he was identified as "a person." Only those adjectives which Sommer and Killian found to be most discriminating were used in the analysis of data. The adjectives were rated on a five point scale ranging from +2 (very complimentary) through 0 (indeterminate) to -2 (very uncomplimentary) but for purposes of analysis the scale was collapsed into three categories and a 3x2 contingency table formed. The rows consisted of subjects who rated the adjective as +2 or +1; those who rated it 0; and those who rated it -2 or -1. The columns were, of course, "Negro" and "Person." The results are presented in Table 1.

TABLE 1

CHI SQUARE AND PROBABILITIES FOR ADJECTIVE RATINGS OF "NEGRO" AND "PERSON"*

Adjective	The University of Michigan**		North Carolina State College***	
	Chi Square	Probability	Chi Square	Probability
Meek	5.24	< .10	.35	< .90
Quiet	3.96	< .20	2.18	< .50
Obedient	1.46	< .50	9.09	< .02
Conservative	.51	< .80	.80	< .70
Dominant	3.26	< .20	2.90	< .30
Carefree	2.95	< .30	3.03	< .30
Forward	3.74	< .20	.76	< .70
Passionate	4.69	< .10	.42	< .90

*All probabilities based on a two-tailed test, 2 degrees of freedom.

**N = 56

***N = 48

Only one of the probabilities in Table 1 is less than .05, and it did not occur in The University of Michigan sample. These results failed to support adequately any distinctive characterization of Negroes by the samples of the study in terms of the adjective check list. The adjective check list thus was not employed further in the studies reported here.

2. THE TRIAD ADJECTIVE FORM

While the analysis of the adjective check list data seemed to indicate that the check list would not be a helpful instrument, the possibility remained that some of the adjectives from it could be used to obtain underlying dimensions which might be used to characterize minority groups. The Unfolding Technique (Coombs, 1950) was suggested for this type of analysis.

In the Fall of 1960 the eight adjectives of Table 1 were presented in combinations of three to 200 students of introductory psychology at The University of Michigan (see forms in Appendix B). Half of the forms distributed at random asked subjects to indicate the adjective of a triad which would be most complimentary and the adjective of the triad which would be the least complimentary for a Negro. The other half of the forms were identical except that "a person" was substituted for "a Negro." The frequencies with which a subject ranked an adjective first, second, and third were tabulated. From these tabulations a summed score was obtained for the preference for each adjective by weighing the ordered positions of first, second, and thirds by 2, 1, and 0 respectively. And individual's ordered preference for the adjectives was then defined by the magnitude of each adjective's score. Thus an order of the eight adjectives was obtained for all subjects. Application of the Unfolding Technique failed to support the existence of an underlying dimension in the characterization of either "a Negro" or "a person." The Triad Adjective Form thus was not employed further in the studies reported here.

3. THE SYLLOGISTIC REASONING TEST

Thistlethwaite (1950) found that subjects from Southern states (i.e., subjects who were assumed to be more ethnocentric than those from the North) made more errors in reasoning on the ethnocentric items of a syllogistic reasoning test than did a group from Northern states. He used 36 syllogisms "containing statements and especially conclusions about which highly ethnocentric respondents have strong attitudes and beliefs" and 36 matched but neutral content syllogisms. These 72 syllogisms were administered to 100 students of introductory psychology at The University of Michigan in the Fall of 1960 (see Appendix C).

A frequency was obtained for each matched pair of syllogisms for the number of subjects who answered the neutral syllogism correctly but missed the paired content syllogism. This frequency was used to order the pairs of syllogisms in terms of the liklihood that prejudice might have influenced reasoning. Some of the content items were on the subject of women. Since this group was not of concern in the studies reported here these items and their matched opposites were not considered further. Of the remaining pairs, however, the five with the highest frequency of biased responses were chosen for use in further studies (see Appendix D).

In the Spring of 1961 the 10 selected syllogisms were given during the first four weeks and again during the final two weeks to 104 students of introductory psychology at The University of Michigan. Each individual was scored on the pre-test and the post-test in terms of the number of times he answered the neutral content item correctly but its matched content item incorrectly. Pre- and post-test means were computed for the men and women of the study. These results are given in Table 2. The results gave no indication that significant changes took place in reasoning bias during the period between the pre- and post-test. The syllogistic reasoning test was not employed further in the studies reported here.

TABLE 2

PRE-TEST MEANS AND MEAN CHANGES IN SYLLOGISTIC REASONING

BEFORE AND AFTER AN INTRODUCTORY COURSE IN PSYCHOLOGY

	N	$\bar{X}$*	$\bar{D}$**	Probability of t by 2-Tail Test
Men	45	6.3	-.07	not significant
Women	59	6.1	-.24	not significant

*A constant, 5, was added to each score to avoid negative numbers.

**A negative mean change ($\bar{D}$) indicates an increase in score (i.e., a change toward more prejudiced reasoning).

4. THE F-SCALE

In the Spring Semester of 1961 the 29 item F-scale (see Appendix E) was administered as a pre-test and post-test to 76 students of introductory psychology at The University of Michigan. The administration of these forms was part of that reported for the syllogistic reasoning test above with the items of that test and the F-scale constituting alternate forms distributed at random in recitation sections. Pre- and post-test means were computed for the men and women of the study. These results are given in Table 3. These results do not support the hypothesis that significant changes took place in authoritarianism during the period between the pre- and post-test.

TABLE 3

PRE-TEST MEANS AND MEAN CHANGES IN F-SCALE AFTER AN INTRODUCTORY COURSE IN PSYCHOLOGY

	N	$\bar{X}$	$\bar{D}$*	Probability of t by 2-Tail Test
Men	41	79.8	1.46	not significant
Women	35	81.4	.37	not significant

*A negative mean change ($\bar{D}$) indicates an increase in score (i.e., a change toward more prejudiced reasoning).

5. ROKEACH'S DOGMATISM SCALE

Rokeach (1960) in an attempt to measure the "openness or closedness of belief systems" developed his Dogmatism Scale. The shorter 40-item form (see Appendix F) was administered to 200 students of introductory psychology in the Fall of 1960. This administration was undertaken to construct the Psychological Dogmatism Scale to correlate with this scale (see Psychological Dogmatism Scale, below).

In the Spring Semester of 1961 the 40-item Dogmatism Scale was administered as a pre- and post-test in alternate forms to those of the syllogistic reasoning test and the F-scale described above. Pre- and post-test means were computed for the men and women of the study. These results are given in Table 4. These results do not support the hypothesis that significant changes took place in dogmatism during the period between the pre- and post-test.

TABLE 4

PRE-TEST MEANS AND MEAN CHANGES IN ROKEACH DOGMATISM BEFORE AND AFTER AN INTRODUCTORY COURSE IN PSYCHOLOGY

	N	$\bar{X}$	$\bar{D}$*	Probability of t by 2-Tail Test
Men	34	133.6	-2.4	not significant
Women	51	134.2	.6	not significant

*A negative mean change ($\bar{D}$) indicates an increase in score (i.e., a change toward more prejudiced reasoning).

6. THE PSYCHOLOGICAL DOGMATISM SCALE

In the Fall of 1960 the principles outlined by Rokeach (1960) for the construction of his Dogmatism scale were used to construct parallel items referring to psychological subject matter. Sixteen of these items (see Appendix G) were administered to 200 students of introductory psychology along with the Rokeach Dogmatism items as indicated above. Item analysis was performed to select 10 (see indicated items in Appendix E) of these 16 items to discriminate the top and bottom third of the distribution of Rokeach Dogmatism scores.

The 10 selected psychological dogmatism scale items were administered to 85 students of introductory psychology at The University of Michigan in the Spring Semester of 1961. These items were placed in the same form as the Rokeach Dogmatism items at the end of that form and were given thus as a pre- and post-test along with that scale. The mean scores were computed for the men and women of the study and these results are given in Table 5. These results do not support the hypothesis that significant changes took place in dogmatism during the period between the pre- and post-test.

TABLE 5

PRE-TEST MEANS AND MEAN CHANGES IN PSYCHOLOGICAL DOGMATISM BEFORE AND AFTER AN INTRODUCTORY COURSE IN PSYCHOLOGY

	N	$\overline{X}$	$\overline{D}$*	Probability of t by 2-Tail Test
Men	34	32.9	.94	not significant
Women	51	31.7	.86	not significant

*A negative mean change ($\overline{D}$) indicates an increase in score (i.e., a change toward more prejudiced reasoning).

7. THE SEMANTIC DIFFERENTIAL SCALES

Osgood's semantic differential provides a technique for evaluating any sort of object on a number of standard scales. This technique was employed in the Spring of 1961 for the evaluation of several ethnic-occupational groups. Twenty-five 7-point bipolar scales were used as the set of standard scales. These were selected (see Appendix H) on the basis of research conducted by O'Dowd and Beardslee (1960) on college student images of occupational groups. A total sample of 630 introductory students were asked to characterize the Jewish College Professor, the Oriental-American College Professor, the Negro College Professor, and the Professor of Psychology on these scales. Of the

630 subjects, 110 were also asked to characterize the Jewish Business Executive, the Oriental-American Business Executive, the Negro Business Executive, the Jewish Construction Worker, the Oriental-American Construction Worker, and the Negro Construction Worker. Another subgroup of 105 was asked to characterize the Psychiatrist and the Social Worker-Psychologist. The remaining individuals of the 630 total sample received either the Syllogistic Reasoning Test, the F-scale, or the Dogmatism Scales as described above. The Semantic Differential Scales were thus administered as a pre-test and post-test as described for these alternative forms which were used in conjunction with the standard scales.

A derogatory score was obtained for each ethnic-occupational group by adding the item responses. Responses on reversed scales were corrected to increase toward the arbitrarily designated negative end of the scale. (This designation corresponded with rare exceptions to the end of the scale avoided by subjects as indicated by scale means.) The mean derogatory scores were computed and are given in Table 6. The derogatory scores obtained in the study permitted more precise comparison of each occupational-ethnic group. The means for the Construction Worker when compared with those for the other groups are strikingly distinct. The finding suggests strongly that occupational-social level considerations overrode any generalized characterization of ethnic groups by the subjects of the study.

Two significant mean changes occurred in a negative direction (i.e., toward higher derogatory scores). These changes for the Negro Business Executive and the Professor of Psychology were not anticipated in the present study. The finding demonstrates, however, that the effects obtained were not entirely due to a pre-post-test difference in response set, i.e., there is no indication that subjects were generally more critical for the post-test. The opposite directed changes suggest a more meaningful hypothesis in terms of relating group prejudice to perceptual processes. In particular, this hypothesis would hold that the Professor of Psychology is characterized less favorably and the Construction Worker more favorably because of a leveling effect in discriminating between groups. The Construction Worker starts with a high mean derogatory score while the Professor of Psychology begins with a low one. The hypothesis would suggest changes toward some general mean.

This explanation of the derogatory score changes for the Professor of Psychology and Construction Worker was somewhat contradictory to an initial hypothesis of the present study. It was anticipated that students would become more discriminating in characterizing ethnic group members more negatively at low occupational levels and more positively at high occupational levels. This hypothesis was tested in terms of the range of occupational derogatory scores within each ethnic group. This range was referred to as a differentiation score. The initial means and mean changes are given by sex in Table 7. Each mean change in differentiation score was significant for all three ethnic groups but not for the psychologists. The changes for the

TABLE 6

PRE- AND POST-TEST DEROGATORY SCORE MEANS AND STANDARD DEVIATIONS

Ethnic Category	Test	College Professor			Business Executive			Construction Worker		
		$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N
Jewish	Pre	78.78	(20.00)	630	78.18	(17.78)	110	102.95	(21.14)	110
								**		
	Post	76.25	(17.64)		79.56	(18.10)		97.87	(18.30)	
Oriental	Pre	73.20	(16.85)	630	70.78	(14.87)	110	94.46	(21.31)	110
								**		
	Post	72.69	(14.80)		73.85	(15.49)		91.19	(17.46)	
Negro	Pre	72.63	(18.52)	630	74.00	(17.75)	110	112.63	(20.05)	110
					**			**		
	Post	73.84	(16.40)		79.58	(16.31)		105.78	(19.34)	
		Professor of Psychology			Psychiatrist			Social Worker		
Unspecified	Pre	71.70	(14.21)	630	70.34	(19.47)	105	69.23	(21.00)	105
		**								
	Post	74.50	(15.10)		68.61	(15.65)		69.43	(16.22)	

**Significant difference in pre-post-test mean by t-test at .01 level of significance, 2-tail test.

ethnic groups with respect to derogatory scores is clearly contradicted. This finding thus supports the argument that students became less discriminating within ethnic groups through the period of taking the course in introductory psychology.

TABLE 7

MEAN PRE-TEST DIFFERENTIATION SCORES AND

MEAN PRE-POST-TEST CHANGES**

		Jewish		Oriental		Negro		Psychologist	
		$\bar{X}$	$\bar{D}$ t	$\bar{X}$	$\bar{D}$ t	$\bar{X}$	$\bar{D}$ t	$\bar{X}$	$\bar{D}$ t
Men	N=41	24.5	-5.29*	22.6	-4.49*	31.6	-11.50*	13.6	-1.14
Women	N=55	20.4	-3.79*	19.5	-3.62*	28.6	- 6.85*	11.8	.54

*Significant mean change at .05 level of significance by 2-tail t-test.

**A negative mean change indicates a change toward a smaller score (i.e., less differentiation).

In connection with the possibility of a general movement toward a norm which these findings seem to indicate, certain considerations are significant regarding the failure to find significant pre-post-test changes for other forms. It may be noted that the initial mean F-scale item response (in the study reported above) is 1.3 or between the responses "I disagree very much" and "I disagree on the whole." In short, the students came into the classroom very low, on authoritarianism. If this is the norm toward which the class situation would move them, it is not surprising that little or no change should be observed. Similarly, students began with a mean item score of 3.3 on Rokeach's Dogmatism scale and of 3.2 on the Psychological Dogmatism scale or between the responses "I disagree a little" and "I agree a little." If an unwillingness to take an extreme position on these items was the norm toward which the class situation would move them, then the absence of change is again not surprising. This explanation is clearly upheld in the case of some items of the Student Opinion Scale.

8. THE STUDENT OPINION SCALES

Newcomb (1961) developed a number of opinion scales which tap issues salient in the college setting. Newcomb's scales on the issues of sex and family, religion, racial problems, and politics were employed in the Spring of 1963 (see Appendix I). To these four scales were added 10 items of the same format but related to assumptions regarding the nature of man and of psychology or psychologists (see Appendix I). These scales were administered to introductory psychology students at The University of Michigan in the Spring Semester of 1963 as a pre-test and again as a post-test. Usable pre- and post-test data were obtained from 255 students. The pre-test was administered at special testing sessions at the beginning of the semester along with other assessment devices used to measure other personality variables. The post-test was administered in class during the last two weeks of the semester along with a course evaluation form.

The items designed to produce characterisations of psychology, psychologists, and man are given in Table 8. Table 8 also gives the percent of the 255 subjects which indicated each alternative as "most acceptable." For example, for the first item, "To what extent is psychology a science?", 80% of the respondents were split equally between alternatives d and e on the pre-test. For the post-test, however, 47% found most acceptable alternative d, "Psychology is basically a science but lacks scientific rigor in some areas."

When asked to what extent psychology should be employed to influence human behavior, Item 2: 44% on the pre-test indicated alternative e, that it should find extensive use. A majority of 63% felt that psychologists should have the same influence as any other group in society, Item 3. No clearcut alternative c of the fifth item was clearly a popular response with 60% placing psychology in the same category as other sciences with respect to potential abuses.

Similarly regarding the items of the second scale on Man, 40% initially see man as about equally determined and free. Or respondents were split about equally on the second item, however, regarding similarity or dissimilarity of man and the lower animals; 40% expressed caution in extending friendship in the third item. When asked whether man is basically good or bad (Item 4) almost a third of the respondents indicated he is good while another third said he was equally good or bad or neutral. And on Item 5 a majority of the respondents indicated they, as a proctor, would not report their friend for cheating.

Aside from the question of the most popular responses the study was concerned with changes which might have occurred during the introductory course. To test for such changes we obtained pre- and post-test means for each item after arbitrarily assigning the numerical values 1 through 5 to the item alternatives a through e respectively. These means are given in Table 9. Table 9 also shows t-test results for the differences between pre- and post-test means for each item. From Table 9 it appears that most of the items on "Psychology and Science" did not produce significant mean changes. The exception is the second item on which the post-test mean indicated somewhat more support for extensive use of psychology to build a better life for mankind. For the items on human nature, however, 3 of the 5 are toward perception of greater determinism in human behavior, perception of less distinction between man and the lower animals, and perception of man as less predisposed to be good.

TABLE 8

PERCENT OF 255 SUBJECTS INDICATING EACH ALTERNATIVE AS: MOST ACCEPTABLE, PRE-TEST (LEFT COLUMN) AND POST-TEST (RIGHT COLUMN)

I. Psychology and Science

Pre* Post*

1. To what extent is psychology a science?
0 1 a. Psychology is not a science at all.
5 4 b. A small part of the field of psychology is scientific.
16 11 c. Psychology has about equal portions of scientific material and non-scientific material.
40 47 d. Psychology is basically a science but lacks scientific rigor in some areas.
40 33 e. Psychology is as much a science as any natural science.

2. To what extent should psychology be employed to influence human behavior?
6 2 a. It should never be consciously used to influence anyone.
11 9 b. It should be used to influence only one's self.
16 16 c. It should be used to influence others in a few restricted ways.
22 22 d. It should be used to influence others as much as one considers desirable.
44 45 e. It should be used extensively to guide people in building a better life for mankind in general.

3. What should be the role of psychologists in society?
1 2 a. They should develop their field but leave its use to others.
21 12 b. They should be active in just a few areas such as education and mental health.
63 69 c. They should have the same influence in society as any other occupational group.
14 13 d. They should have an important role in deciding how society should be run.
0 1 e. They should be the most influential group in governing society.

4. To what extent should courses in psychology be required?
30 30 a. Psychology should be an elective in high school and college.
30 25 b. It should be an elective in high school but one course should be required of certain majors in college.
25 29 c. It should be an elective in high school but one course should be required of all college students.
11 11 d. A course should be required of all high school and another course should be required of all college students.

TABLE 8 (Continued)

Pre*	Post*	
2	1	e. It should be a required course starting in grade school and continuing at least through one course in high school and one course in college.
		5. To what extent should people be concerned about the uses of psychology?
7	9	a. There is nothing to be concerned about; psychology is quite harmless.
20	15	b. Psychology offers very little to worry about if people in general do not lose their heads.
60	64	c. Psychology, like any body of knowledge, provides its possessor with power which can be abused and people must be vigilant in guarding against this abuse.
5	5	d. Psychology is especially dangerous among the bodies of knowledge because it can provide considerable power over other human beings.
4	2	e. The dangers of psychology are far more grave than is generally recognized; psychology can potentially provide dangerous forms of inhumanity.

II. Man **

Pre*	Post*	
		1. To what extent is human behavior determined?
8	5	a. Man is a free agent depending only on his will to be free.
27	22	b. Man is generally free within a few limitations of the natural world.
40	27	c. Man is determined in certain respects but may exercise free will in about as many other respects.
17	28	d. Man is generally determined but can exercise free will in a few instances.
5	12	e. Man is entirely determined by laws of the natural world.
		2. To what extent is man apart from the lower animals?
20	20	a. Man is basically an animal differing from the lower animals only in terms of complexity.
27	35	b. There are a few characteristics of man which set him apart from the lower animals but for the most part man is only a more complex animal.
17	14	c. The lower animal and uniquely human elements are represented in man in about equal proportion depending on the person.
29	23	d. Man is unlike the lower animals in more respects than he is like the lower animals.
4	2	e. The characteristics of man are such that analyses of animal behavior have little or no relevance to the study of human behavior.

TABLE 8 (Concluded)

Pre*	Post*	
		3. To what extent should a person try to become close friends with others?
3	2	a. Be self-sufficient and don't form close ties with anyone; one doesn't get hurt that way.
40	37	b. Form close ties with only a few people who are really understanding and can be trusted.
14	20	c. Become close friends with anyone you trust, a lot of people can be trusted but a lot cannot.
10	12	d. Try to become close friends with all the people you know, most people will be loyal friends if they know they are trusted, through a few may take advantage of such trust.
31	24	e. Let people know you trust them and want to be close friends with them; they will respond in kind.
		4. Does human nature predispose man to be good or evil?
31	25	a. Man is basically good.
22	20	b. Man is predominantly good.
34	38	c. Man is neutral or equally good and evil.
4	5	d. Man is predominantly evil.
4	6	e. Man is basically evil.
		5. Suppose a man is acting as a student proctor in an examination for some large, introductory course. He sees that his roommate and best friend is cheating in the exam. The roommate is a hard-working though not a brilliant student, and desperately needs a good grade in this course. The decision as to what to do about it is entirely in the proctor's hands, for no one else sees the cheating. What should the proctor do?
14	14	a. Take away his exam book and report him for cheating.
26	27	b. Prevent him from cheating further, let him finish the exam, and report him for cheating.
17	12	c. If he can be led to withdraw from the exam on some excuse, treat him as absent; otherwise report him for cheating.
38	40	d. Prevent further cheating but let him finish the exam, and not report him for cheating.
3	2	e. Act as if nothing had happened, and not report him for cheating.

---

*Item columns do not sum to 100 because some subjects declined to answer some items.

**N=248 subjects for these items. The 7 subjects dropped, did not complete this part or did so improperly.

TABLE 9

DIFFERENCES BETWEEN MEANS, PRE- MINUS POST-TEST MEANS*

Item	Pre-Test Mean	$\bar{x}$ Pre- $\bar{x}$ Post	$\sigma_{\bar{x}_1-\bar{x}_2}$	$r_{pre-post}$	t	Sign Level 2-Tail**
I_1	4.14	.04	.071	.183	.56	$p > .25$
I_2	3.85	-.18	.076	.522	-2.37	$p < .02$
I_3	2.90	-.08	.047	.312	1.70	$p < .10$
I_4	2.21	-.05	.078	.371	- .64	$p > .25$
I_5	2.79	.04	.076	.120	.52	$p > .25$
II_1	2.82	-.37	.068	.597	5.44	$p < .002$
II_2	2.68	.20	.096	.234	2.08	$p < .05$
II_3	3.24	.04	.093	.401	.43	$p > .25$
II_4	2.27	-.27	.095	.290	2.84	$p < .01$
II_5	2.90	.00	.078	.541	.00	$p > .25$

*Scale values of responses arbitrarily assigned: a = 1, b = 2, c = 3, d = 4, e = 5.

**NB: The 5 significant t's are not independent tests.

Certain patterns of changes in the psychology and psychologists questions suggested that the use of mean differences was not an appropriate form of analysis. In particular, movement toward the center alternatives from both ends were evident in some of these items. A different analysis revealed that in spite of the lack of significant differences in means, certain changes did take place in Items 1, 2, 3, and 5 of those related to "Psychology and Science." These changes displayed a pattern of conformity. This conformity took place toward the norm defined by the modal response of the pre-test as given in Table 10. The actual percentage of Ss who conform by giving this response on the post-test is compared with the expected percentage if the respondents had answered at random on the post-test. Table 10 gives the Z's computed from the estimated and actual percentages. The significance of each Z is less than .05 in four of the five cases. The exception is Item 4 for which the number staying with their original response or changing to alternative c results in a significantly smaller percentage of conformists than would be expected by chance. Conformity in the case of the other four items took toward the view that psychology is basically a science but...(Item 1, alternative d), extensive use of psychology is desirable to build a better life for mankind (Item 2, alternative e) psychologists should have the same influence as other groups (Item 3, alternative c).

There were 98 males and 157 females in the sample used. Attempts to find sex differences in conformity were unsuccessful. No relationship could be found between grade in course and conformity. It would seem that the attitudinal changes occurring were quite general in terms of sex and grade.

TABLE 10

PERCENT OF RESPONDENTS NOT GIVING MODAL RESPONSE ON PRE-TEST WHO GAVE THAT RESPONSE OF POST-TEST

Item	Modal Response Pre-Test	No. Not Giving Pre-Test Modal Response	Percent Conforming to Pre-Test Mode on Post-Test	Expected Percentage	σ_p	Z	Sign Level 1-Tail*
I_1	d or e	52	51.9	40	.068	1.75	$p < .05$
I_2	e	143	30.8	20	.033	3.27	$p < .001$
I_3	e	94	53.2	20	.041	8.10	$p < .001$
I_4	a or b	99	31.3	40	.049	-1.77	wrong direction
I_5	c	102	45.1	20	.040	6.27	$p < .001$

*NB: The significance tests are not independent.

The results reported here suggest that similar analysis of the other scales of the Student Opinion Scales should prove fruitful. This analysis has not been completed as of this writing.

9. CONCLUSIONS

A number of attitudinal measures were studied in the seris of studies reported here. The standard measures such as the F-scale, Rokeach's Dogmatism Scale, syllogistic reasoning tests, and adjective check lists were conspicuously unhelpful. The results provide further support for Jacob (1957) when he concludes that either significant attitude changes do not generally take place during college courses or the instruments to measure them are not adequately developed. By contrast to these discouraging results more issue-specific, perhaps more salient, items reveal certain clearcut changes during the introductory course in psychology. Student characterizations of psychologists and psychology, in particular, appear to change toward some class norm. And student characterization of human nature appears to move in a consistently unflattering direction. The results suggest that measures of attitude change may need to be made in terms of some norm which may or may not be on the ends of one of the standard scales. Thus, convergence on the norm may not be reflected in an overall mean change. Further study is necessary to establish the significance of this type of change and whether it is due to the experience of the introductory course in psychology or to the semester's college experience in general.

References

Coombs, C. H. Psychological scaling without a unit of measurement. _Psychol. Rev._, 1950, 57, 145-158.

Jacob, P. E. _Changing Values in College_. New York: Harper, 1957.

Newcomb, T. E. _The Acquaintance Process_. New York: Holt, Rinehart and Winston, 1961.

O'Dowd, D. D. and Beardslee, D. C. _College Student Images of a Selected Group of Occupations_. Middletown, Conn.,: Authors, 1960.

Rokeach, M. _The Open and Closed Mind_. New York: Basic Books, 1960.

Thistlethwaite, C. Attitude and structure as factors in the distortion of reasoning. _J. abnorm. soc. Psychol._, 1958, 57, 37-40.

Sommer, R. and Killian, L. Areas of value difference: A method for investigation. _J. soc. Psychol._, 1954, 39, 227-235.

APPENDIX A

Directions: If you heard a Negro referred to in terms of each of the following adjectives, would it be complimentary or uncomplimentary to him? Indicate your answer by inserting a number to the left of each adjective according to the following scale:

if very complimentary mark	+2
complimentary mark	+1
indeterminate mark	0
uncomplimentary mark	-1
very uncomplimentary mark	-2

_____sociable
_____gentle
_____enthusiastic
_____proud
_____passionate
_____loyal
_____submissive
_____talkative
_____eager
_____carefree
_____kind
_____practical
_____religious
_____critical
_____strict
_____serious
_____unsympathetic
_____helpful
_____shy
_____slow
_____jovial
_____trusting
_____faithful

_____shrewd
_____well-liked
_____independent
_____ambitious
_____neat
_____modern
_____liberal
_____thrifty
_____obedient
_____quiet
_____neighborly
_____careful
_____jealous
_____bitter
_____graceful
_____calm
_____meek
_____stubborn
_____aggressive
_____conservative
_____modest
_____forward
_____timid

_____happy-go-lucky
_____athletic
_____witty
_____progressive
_____elegant
_____nationalistic
_____respectable
_____musical
_____quick-tempered
_____cautious
_____clever
_____quick
_____loud
_____childish
_____patient
_____dependent
_____senstive
_____traditional
_____good-natured
_____persistent
_____dominant
_____original
_____sympathetic

Directions: If you heard a person referred to in terms of each of the following adjectives, would it be complimentary or uncomplimentary to him? Indicate your answer by inserting a number to the left of each adjective according to the following scale:

if very complimentary mark +2
complimentary mark +1
indeterminate mark 0
uncomplimentary mark -1
very uncomplimentary mark -2

_____sociable
_____gentle
_____enthusiastic
_____proud
_____passionate
_____loyal
_____submissive
_____talkative
_____eager
_____carefree
_____kind
_____practical
_____religious
_____critical
_____strict
_____serious
_____unsympathetic
_____helpful
_____shy
_____slow
_____jovial
_____trusting
_____faithful

_____shrewd
_____well-liked
_____independent
_____ambitious
_____neat
_____modern
_____liberal
_____thrifty
_____obedient
_____quiet
_____neighborly
_____careful
_____jealous
_____bitter
_____graceful
_____calm
_____meek
_____stubborn
_____aggressive
_____conservative
_____modest
_____forward
_____timid

_____happy-go-lucky
_____athletic
_____witty
_____progressive
_____elegant
_____nationalistic
_____respectable
_____musical
_____quick-tempered
_____cautious
_____clever
_____quick
_____loud
_____childish
_____patient
_____dependent
_____sensitive
_____traditional
_____good-natured
_____persistent
_____dominant
_____original
_____sympathetic

APPENDIX B

ADJECTIVE CHECK LIST

This list contains 56 sets of three adjectives each. Within each set of three adjectives you are to choose the one which would be most complimentary if used to describe a Negro and the one which would be least complimentary. On the accompanying answer sheet mark a for the number of the adjective in each set of three which would be most complimentary in describing a Negro; mark b for the adjective which would be least complimentary; mark c for the remaining one of the three adjectives.

Illustrated below are some of the ways your choices for the first set could be marked. This is only an example—you could have other choices.

	a	b	c	d	e		a	b	c	d	e		a	b	c	d	e
1.						1.						1.					
2.						2.						2.					
3.						3.						3.					

etc.

DO NOT MARK ON THIS BOOKLET

ADJECTIVE CHECK LIST

1. meek
2. obedient
3. forward

4. quiet
5. conservative
6. obedient

7. obedient
8. carefree
9. quiet

10. dominant
11. forward
12. obedient

13. conservative
14. obedient
15. dominant

16. carefree
17. quiet
18. conservative

19. obedient
20. meek
21. carefree

22. conservative
23. obedient
24. meek

25. conservative
26. carefree
27. passionate

28. quiet
29. meek
30. conservative

31. meek
32. quiet
33. obedient

34. forward
35. conservative
36. passionate

37. forward
38. dominant
39. conservative

40. passionate
41. forward
42. dominant

43. dominant
44. forward
45. quiet

46. forward
47. quiet
48. conservative

49. forward
50. quiet
51. meek

52. forward
53. quiet
54. obedient

55. dominant
56. carefree
57. conservative

58. dominant
59. meek
60. forward

61. obedient
62. passionate
63. forward

64. meek
65. conservative
66. forward

67. quiet
68. carefree
69. forward

70. forward
71. dominant
72. carefree

73. meek
74. obedient
75. dominant

76. meek
77. conservative
78. passionate

79. forward
80. conservative
81. obedient

82. passionate
83. meek
84. dominant

85. quiet
86. conservative
87. passionate

88. carefree
89. conservative
90. obedient

91. dominant
92. quiet
93. carefree

94. carefree
95. dominant
96. meek

97. passionate
98. meek
99. quiet

Adjective Check List, continued

100. quiet
101. passionate
102. forward

103. conservative
104. obedient
105. passionate

106. conservative
107. dominant
108. meek

109. passionate
110. forward
111. meek

112. conservative
113. quiet
114. dominant

115 dominant
116. passionate
117. quiet

118. meek
119. passionate
120. carefree

121. quiet
122. meek
123. carefree

124. passionate
125. obedient
126. quiet

127. obedient
128. dominant
129. passionate

130. carefree
131. passionate
132. dominant

133. quiet
134. carefree
135. passionate

136. carefree
137. forward
138. meek

139. obedient
140. passionate
141. meek

142 dominant
143. obedient
144. quiet

145. passionate
146. forward
147. carefree

148. passionate
149. dominant
150. conservative

151. carefree
152. meek
153. conservative

154. obedient
155. carefree
156. forward

157. carefree
158. passionate
159. obedient

160. conservative
161. forward
162. carefree

163. obedient
164. carefree
165. dominant

166. meek
167. dominant
168. quiet

APPENDIX C

NAME________________________________ STARTED______________

HAVE YOU HAD A COURSE IN LOGIC? ____________ FINISHED______________
(yes or no)

INSTRUCTOR______________________________ SECTION NUMBER__________

Test of Reasoning Ability—Form I

Instructions

The following test contains a series of items designed to test reasoning ability. Some of the items will be very easy. Some will be more difficult. In each case a statement is given, followed by a conclusion which begins, "Therefore..."

You are to determine whether the conclusion follows from the given statement or statements.

If the conclusion follows from, or is implied in, the given statement(s) mark the item true by circling the "T."

If the conclusion does not follow from, or is not implied in, the given statement(s) mark the item false by circling the "F."

Example: Given: If the witness told the truth, then the defendant is guilty. The witness told the truth.

[✓ T] F Therefore: The defendant is guilty.

You will be timed, so proceed rapidly after the starting signal is given. As soon as you complete the test bring it to the instructor and he will record the time finished on your paper.

At the signal "Begin work" proceed immediately to the test. Do not omit any items. Make a judgment on each item, whether you are sure or not. Do not turn back. Keep going until you finish.

DO NOT TURN THIS PAGE BEFORE THE SIGNAL IS GIVEN

1. Given: Pueblo Indians who are married are likely to ridicule and make fun of visitors. All persons who are likely to ridicule and make fun of visitors should be considered emotionally insecure.
   T F Therefore: Pueblo Indians who are married should be considered emotionally insecure.

2. Given: If skilled work does not require speed then Negroes are best fitted for skilled work. If it is not the case that Negroes are best fitted for positions of responsibility, then dependability is unimportant. If skilled work requires speed then dependability is not unimportant.
   T F Therefore: Negroes are best fitted for skilled work or for positions of responsibility.

3. Given: If the Wickersham report is correct, then there is a high proportion of illiteracy in the criminal group. If the Wickersham report is correct then there is a high frequency of divorce among the parents of criminals.
   T F Therefore: If there is a high proportion of illiteracy in the criminal group, then there is a high frequency of divorce among parents of criminals.

4. Given: If the proofreaders were responsible for the errors of fact in the new book then they should be admonished to be more careful. The proofreaders should be admonished to be more careful.
   T F Therefore: The proofreaders were responsible for the errors of fact in the new book.

5. Given: If Jews are iconoclastic or if they are critical of others then they are always unpopular. If Jews are not critical of others then they should be encouraged to join Gentile fraternities. Jews are not always unpopular.
   T F Therefore: Jews should be encouraged to join Gentile fraternities.

6. Given: If this is a desirable neighborhood, then it is restricted to Gentiles. This is not a desirable neighborhood.
   T F Therefore: It is not restricted to Gentiles.

7. Given: If the history is not inaccurate then the plight of the Negroes is the fault of bigots. If the plight of the Negroes is not the fault of Fascists then Fascism is of recent origin. If the history is inaccurate then Fascism is not of recent origin.
   T F Therefore: The plight of the Negroes is the fault of bigots or of Fascists.

8. Given: If women are by nature leaders, then women should be encouraged to run for the presidency of the United States. It is not the case that women are by nature leaders.
   T F Therefore: Women should not be encouraged to run for the presidency of the United States.

9. Given: If women are naturally maternal then the proper role of women is to care for the home and children. If women are naturally maternal then women should attend to domestic affairs before getting into politics.
   T F Therefore: If the proper role of women is to care for the home, they should attend to domestic affairs before getting into politics.

1. Given: If this is a desirable neighborhood, then it is close to transportation services. This is not a desirable neighborhood.
   T F Therefore: It is not close to transportation services.

2. Given: If survival of the fittest is a law of nature, then America deserves to exercise leadership over backward nations. If survival of the fittest is a law of nature, then America should not subordinate its interests to those of other nations.
   T F Therefore: If America deserves to exercise leadership over backward nations then it should not subordinate its interests to those of other nations.

3. Given: If the men who built the bridge were not careful then they are responsible. If they are not residents then they were ignorant of the plan. If they were careful then they were not ignorant of the plan.
   T F Therefore: The men who built the bridge are responsible or they are residents.

4. Given: If women are naturally weak and dependent then it is always best for them to bear children. If women are not dependent then they should be encouraged to act independently. It is not always best for women to bear children.
   T F Therefore: Women should be encouraged to act independently.

5. Given: If the basketball team is the best in the conference then other teams often attempt to copy its plays. Other teams often attempt to copy the basketball team's plays.
   T F Therefore: The basketball team is the best in the conference.

6. Given: Negroes who are educated are likely to make distinctive contributions to American culture. All persons who are likely to make distinctive contributions to American culture should be admitted to the better hotels and apartments.
   T F Therefore: Negroes who are educated should be admitted to the better hotels and apartments.

7. Given: If the Arapesh Indians are superior in ability then they deserve the favorable grounds they inhabit. If the Arapesh Indians are superior in ability then they should not compromise with their ideals.
   T F Therefore: If the Arapesh Indians deserve the favorable grounds they inhabit then they should not compromise with their ideals.

8. Given: If Jews have distinctive gestures or peculiar speech habits then they are always distinguishable from Gentiles. If they do not have peculiar speech habits then they are grossly misrepresented by most people. Jews are not always distinguishable from Gentiles.
   T F Therefore: Jews are grossly misrepresented by most people.

9. Given: If America is the most politically mature nation in the world today then other nations often take their lead from the U.S. Other nations do often take their lead from the U.S.
   T F Therefore: America is the most politically mature nation in the world today.

1. Given: If the Dobu Indians are inherently belligerent then their neighbors only rarely have peaceful contact with them. Neighbors of the Dobu Indians only rarely have peaceful contact with them.
   T F Therefore: The Dobu Indians are inherently belligerent.

2. Given: If the average I.Q. of Negroes is as high as that of Whites then Negroes are the equal of Whites in intelligence. The average I.Q. of Negroes is not as high as that of Whites.
   T F Therefore: Negroes are not the equals of Whites in intelligence.

3. Given: If women are not innately tender-minded then they ought to have careers. If women ought not to be politically active then women lack political ability. If women are innately tender-minded they do not lack political ability.
   T F Therefore: Women ought to have careers or ought to be politically active.

4. Given: Businessmen who employ many people prefer workers who are loyal. Anyone who prefers workers who are loyal should hire as large a percentage of Jews as possible.
   T F Therefore: Businessmen who employ many people should hire as large a percentage of Jews as possible.

5. Given: If conscientious objectors are always stubborn or irresponsible then their judgments are always impressive. If conscientious objectors are not always irresponsible then they ought to be free to disobey their government's laws. The judgments of conscientious objectors are not always impressive.
   T F Therefore: Conscientious objectors ought to be free to disobey their government's laws.

6. Given: If the game is crucial then a large attendance is desirable. If the game is crucial then it is a mistake for the team members not to practice today.
   T F Therefore: If a large attendance is desirable then it is a mistake for the team members not to practice today.

7. Given: If the Negro is by nature lazy and superstitious then he can usually profit little from higher education. The Negro can usually profit little from higher education.
   T F Therefore: The Negro is by nature lazy and superstitious.

8. Given: If the Iroquois have distinctive mental or behavioral characteristics then they are like the Arapesh Indians. If the Iroquois do not have distinctive behavioral characteristics then they are hardly distinguishable from the Zuni. The Iroquois are not like the Arapesh Indians.
   T F Therefore: The Iroquois are hardly distinguishable from Zuni.

9. Given: Rugby players who are active have unusual endurance. All persons who have unusual endurance should be examined periodically to determine whether the heart is being overtaxed.
   T F Therefore: Rugby players who are active should be examined periodically to determine whether the heart is being overtaxed.

1. Given: If the team was overconfident and cocky then they lost the game. The team lost the game.
   T F Therefore: The team was overconfident and cocky.

2. Given: If Negroes are fundamentally primitive or morally weak then they are deserving of better education. If they are not morally weak then they should feel no hesitation about dating White girls who like them. Negroes are not deserving of better education.
   T F Therefore: Negroes should feel no hesitation about dating White girl's who like them.

3. Given: If accidents should be prevented, then traffic lights should be installed at busy intersections. If accidents should be prevented, then it is best to keep one's eyes on the road when driving.
   T F Therefore: If traffic lights should be installed at busy intersections then it is best to keep one's eyes on the road when driving.

4. Given: Students who collect books prefer volumes with good bindings. Anyone who prefers volumes with good bindings should purchase as many first editions as possible.
   T F Therefore: Students who collect books should purchase as many first editions as possible.

5. Given: If radical agitators are responsible for the talk about putting Negroes on the same level as Whites then they deserve to be silenced and prevented from trouble-making. Radical agitators deserve to be silenced and prevented from trouble-making.
   T F Therefore: Radical agitators are responsible for the talk about putting Negroes on the same level as Whites.

6. Given: If a mixture of study and recitation increases efficiency in learning then the subject should avoid prolonged periods of continuous study. A mixture of study and recitation does not increase efficiency in learning.
   T F Therefore: The subject should not avoid prolonged periods of continuous study.

7. Given: If the custom of the potlatch is not harmful then the Crow Indians should preserve it. If it is not the case that the Crow Indians should humiliate each other then reports are correct. If the custom of the potlatch is harmful then reports are not correct.
   T F Therefore: The Crow Indians should preserve the custom of the potlatch or they should humiliate each other.

8. Given: If this is a desirable playground location then there is too much traffic nearby. If this is a desirable playground location then it is convenient.
   T F Therefore: If there is not too much traffic nearby, then the playground location is convenient.

9. Given: If the Jews who crucified Jesus were not repentant, then they were not responsible. If they were not Christians, then they were sympathetic with the plan. If they were repentant then they were not sympathetic with the plan.
   T F Therefore: The Jews who crucified Jesus were not responsible or they were Christians.

1. Given: If the evaluation is not in error then Judaism is fundamentally Christ-like in its teachings. If it is not the case that Judaism is superior to Christianity then the Jewish Bible is not accurate. If the evaluation is in error then the Jewish Bible is accurate.
T F Therefore: Judaism is fundamentally Christ-like in its teachings or it is superior to Christianity.

2. Given: Every Zuni Indian is trained to discipline himself. No one who is trained to discipline himself should submit to the tribal customs without thinking for himself.
T F Therefore: No Zuni Indian should submit to tribal customs without thinking for himself.

3. Given: If the minority opinion is accurate, then Gentiles are just as mercenary and shrewd as Jews. The minority opinion is not accurate.
T F Therefore: Gentiles are not as mercenary and shrewd as Jews.

4. Given: If unnecessary conflicts should be avoided then Negroes should be prevented from having too much contact with Whites. If unnecessary conflicts should be avoided then it is best to keep Negroes in their own districts and schools.
T F Therefore: If Negroes should be prevented from having too much contact with Whites then it is best to keep Negroes in their own districts and schools.

5. Given: If the Iroquois are weak in numbers or of little vitality then it is best for them to cultivate domestic arts. If they are not of little vitality then they should be encouraged to perpetuate their culture. It is not best for the Iroquois to cultivate domestic arts.
T F Therefore: The Iroquois should be encouraged to perpetuate their culture.

6. Given: If the Navajo males tend to form small aristocratic cliques then they are often ridiculed and imitated by Navajo females. The Navajo males are often ridiculed and imitated by Navajo females.
T F Therefore: The Navajo males tend to form small aristocratic cliques.

7. Given: If the Iroquois Indian is of the same stock as the South Sea islander, then the Iroquois migrated to this country from Asia. The Iroquois Indian is not of the same stock as the South Sea islander.
T F Therefore: The Iroquois did not migrate to the country from Asia.

8. Given: The Crow Indians have to do part of their talking by gesturing. All persons who have to do part of their talking by gesturing are unable to talk rapidly.
T F Therefore: Crow Indians are unable to talk rapidly.

9. Given: If armaments are not conducive to peace then America should reduce its Army and Navy. If it is not the case that America should internationalize the secret of the atom bomb then America should destroy all its atom bombs. If armaments are conducive to peace then America should not destroy all its atom bombs.
T F Therefore: America should reduce its Navy and Army or should internationalize the secret of the atom bomb.

1. Given: If the Hopi Indians are not trouble-makers then they ought to have shelters. If the Hopi Indians ought not be left alone then they have suicidal tendencies. If the Hopi Indians are trouble-makers then they do not have suicidal tendencies.
   T F Therefore: The Hopi Indians ought to have shelters or ought to be left alone.

2. Given: If the Pueblo Indians are easy-going or if they are friendly then they are suitable subjects. If they are not friendly then they should be investigated carefully. The Pueblo Indians are not suitable subjects.
   T F Therefore: The Pueblo Indians should be investigated carefully.

3. Given: If this is a nice neighborhood, then there are not too many Jews around. If this is a nice neighborhood, then there is little clannishness.
   T F Therefore: If there are not too many Jews around, then there is little clannishness.

4. Given: If boomerangs are used for self-defense, then the Australian Indians should be encouraged to manufacture them. It is not the case that boomerangs are used for self-defense.
   T F Therefore: Australian Indians should not be encouraged to manufacture boomerangs.

5. Given: American political leaders have to spend much of their time in campaigning. All persons who have to spend much of their time in campaigning are usually unequipped to govern wisely.
   T F Therefore: American political leaders are usually unequipped to govern wisely.

6. Given: If Jewish businessmen tend to stick together and to connive then they are often isolated and excluded from Gentile circles. Jewish businessmen are often isolated and excluded from Gentile circles.
   T F Therefore: Jewish businessmen tend to stick together and to connive.

7. Given: If the Crow Indians do not value speed then they can be expected to work slowly. If it is not the case that the Crow Indians are skillfull then they are handicapped. If the Crow Indians value speed then they are not handicapped.
   T F Therefore: The Crow Indians can be expected to work slowly or they are skillful.

8. Given: If potlatch participants are rough or overly enthusiastic then they always injure each other. If they are not overly enthusiastic then the potlatch is probably no more destructive than other contests. Potlatch participants do not always injure each other.
   T F Therefore: Potlatch is probably no more destructive than other contests.

9. Given: If casual observation is accurate, then the Zuni Indians are modest and unostentatious. Casual observation is not accurate.
   T F Therefore: The Zuni Indians are not modest and unostentatious.

1. Given: If production is important, then peaceful industrial relations are desirable. If production is important, then it is a mistake to have Negroes for foremen and leaders over Whites.
   T F Therefore: If peaceful industrial relations are desirable, then it is a mistake to have Negroes for foremen and leaders over Whites.

2. Given: Women are disposed to spend much of their time at home. All persons who are disposed to stay at home should have some political or civic activity which takes them out into public life.
   T F Therefore: Every woman should have a political or civic activity which takes her out into public life.

3. Given: If the Dobu Indian is best suited for playing a paranoic role then he should never humiliate himself in public. The Dobu Indian should never humiliate himself in public.
   T F Therefore: The Dobu Indian is best suited for playing a paranoic role.

4. Given: If the record is not in error, then the accident was the fault of the driver. If the accident was not the fault of the policeman then the light was red. If the record is in error then the light was not red.
   T F Therefore: The accident was the fault of the driver or it was the fault of the policeman.

5. Given: If mixture with inferior foreign groups increases national vitality then the United States should permit unrestricted immigration into this country. Mixture with inferior foreign groups does not increase national vitality.
   T F Therefore: The U.S. should not permit unrestricted immigration into this country.

6. Given: If the Plains Indians are dangerous or hostile then they need to be disciplined. If the Plains Indians are not hostile then they ought to be free to practice magic. The Plains Indians do not need to be disciplined.
   T F Therefore: The Plains Indians ought to be free to practice magic.

7. Given: If the census report is correct, then there is a high proportion of wealthy men among the Jews. If the report is correct, then there is a high percentage of Jews in the professions.
   T F Therefore: If there is a high proportion of wealthy men among the Jews, then there is a high percentage of Jews in the professions.

8. Given: If the Negroid race is biologically similar to the White race, then the Negroid group represents an advanced and mature race. The Negroid race is not biologically similar to the White race.
   T F Therefore: The Negroid group does not represent an advanced and mature race.

9. Given: Dobu Indians are inclined to practice elaborate magical rituals. All persons who are inclined to practice elaborate magical rituals should be treated with skill and understanding.
   T F Therefore: The Dobu Indian should be treated with skill and understanding.

1. Given: If Negroes are largely uneducated or if they are unskilled then they are treated unfairly. If they are not unskilled then they are treated unfairly. If they are not unskilled then they should be assisted as much as possible in improving their social and economic status. Negroes are not treated unfairly.
   T F Therefore: Negroes should be assisted as much as possible in improving their social and economic status.

2. Given: If the record is not in error then the defendant is guilty. If the defendant is not of age then the parents are to blame. If the record is in error then the parents are not to blame.
   T F Therefore: The defendant is guilty or is of age.

3. Given: Jews who are talented are equipped with an ability to get things done. All persons who are equipped with an ability to get things done should be encouraged to train for administrative and professional positions.
   T F Therefore: Jews who are talented should be encouraged to train for administrative and professional positions.

4. Given: If a woman is best suited for playing feminine or maternal roles then she should never let her career interfere with her marriage. A woman should never let her career interfere with her marriage.
   T F Therefore: A woman is best suited for playing feminine or maternal roles.

5. Given: If the Pueblo culture is typically peaceful then the proper role of the Zuni is to lead an inoffensive life. If the Pueblo culture is typically peaceful then the Zuni Indian does not take marriage too seriously.
   T F Therefore: If the proper role of the Zuni is to lead an inoffensive life then he does not take marriage too seriously.

6. Given: If the average life-span of the Zuni is as long as that of the Melanesians, then the two cultures have the same frequency of suicides. The average life-span of the Zuni is not as long as that of the Melanesians.
   T F Therefore: The two cultures do not have the same frequency of suicides.

7. Given: If the Hopi Indians are superstitious or fanatical then they are a primitive group. If they are not fanatical then they should be taught more efficient agricultural methods. The Hopi Indians are not a primitive group.
   T F Therefore: The Hopi Indians should be taught more efficient agricultural methods.

8. Given: If Jews are conceited and of overbearing pride then they deserve the treatment they get. Jews deserve the treatment they get.
   T F Therefore: Jews are conceited and of overbearing pride.

9. Given: Every Negro should be trained to discipline himself properly. No one who should be trained to discipline himself properly should submit to social conventions without thinking for himself.
   T F Therefore: No Negro should submit to social conventions without thinking for himself.

APPENDIX D

I.D. Number____________________

Name____________________

Section Number____________________

Instructor____________________

Have you had a course
in logic?____________________

Time taken for
Part II____________________

PART II

SYLLOGISTIC REASONING TEST

In this test you will be given a number of syllogisms. Each syllogism will state what is given and what is concluded. If the conclusion follows from what is given, you are to mark the syllogism TRUE. If the conclusion does not follow from what is given, you are to mark the syllogism FALSE.

Instructions

1. Place an unused mark sense card beside the test booklet. Place the green lettered side up with the column of nines on the left, the column of blank spaces on the right. Reading down the left side of the card (ignore small numbers 1 through 80) you will find the rows numbered 1 through 27. Each row corresponds to the possible answers to a question.

2. Indicate whether the ten syllogisms of the test are true or false by using the corresponding rows 1 through 10 on the card. Mark the space numbered 1 if the syllogism is true. Mark the space numbered 2 if the syllogism is false.

3. Use an electrographic pencil. Mark one and only one answer to each item.

4. Time yourself. Do not spend more than 20 minutes. You should guess at unfinished items.

USE ROWS 1 THROUGH 5 ON CARD. MARK 1 IF TRUE, 2 IF FALSE

1. Given: If the Wickersham report is correct, then there is a high proportion of illiteracy in the criminal group. If the Wickersham report is correct then there is a high frequency of divorce among the parents of criminals.

   Therefore: If there is a high proportion of illiteracy in the criminal group, then there is a high frequency of divorce among parents of criminals.

2. Given: If the history is not inaccurate then the plight of the Negroes is the fault of bigots. If the plight of the Negroes is not the fault of Fascists then Fascism is of recent origin. If the history is inaccurate then Fascism is not of recent origin.

   Therefore: The plight of the Negroes is the fault of bigots or of Fascists.

3. Given: If conscientious objectors are always stubborn or irresponsible then their judgments are always impressive. If conscientious objectors are not always irresponsible then they ought to be free to disobey their government's laws. The judgments of conscientious objectors are not always impressive.

   Therefore: Conscientious objectors ought to be free to disobey their government's laws.

4. Given: If the Iroquois have distinctive mental or behavioral characteristics then they are like the Arapesh Indians. If the Iroquois do not have distinctive behavioral characteristics then they are hardly distinguishable from the Zuni.. The Iroquois are not like the Arapesh Indians.

   Therefore: The Iroquois are hardly distinguishable from Zuni..

5. Given: If Negroes are fundamentally primitive or morally weak then they are deserving of better education. If they are not morally weak then they should feel no hesitation about dating White girls who like them. Negroes are not deserving of better education.

   Therefore: Negroes should feel no hesitation about dating White girls who like them.

DO NOT TURN CARD OVER. USE ROWS 6 THROUGH 10

6. Given: If accidents should be prevented, then traffic lights should be installed at busy intersections. If accidents should be prevented, then it is best to keep one's eyes on the road when driving.

   Therefore: If traffic lights should be installed at busy intersections then it is best to keep one's eyes on the road when driving.

7. Given: If unnecessary conflicts should be avoided then Negroes should be prevented from having too much contact with Whites. If unnecessary conflicts should be avoided then it is best to keep Negroes in their own districts and schools.

   Therefore: If Negroes should be prevented from having too much contact with Whites then it is best to keep Negroes in their own districts and schools.

8. Given: If the record is not in error, then the accident was the fault of the driver. If the accident was not the fault of the policeman then the light was red. If the record is in error then the light was not red.

   Therefore: The accident was the fault of the driver or it was the fault of the policeman.

9. Given: If the Plains Indians are dangerous or hostile then they need to be disciplined. If the Plains Indians are not hostile then they ought to be free to practice magic. The Plains Indians do not need to be disciplined.

   Therefore: The Plains Indians ought to be free to practice magic.

10. Given: If the census report is correct, then there is a high proportion of wealthy men among the Jews. If the report is correct, then there is a high percentage of Jews in the professions.

    Therefore: If there is a high proportion of wealthy men among the Jews, then there is a high percentage of Jews in the professions.

* *

In row 26 mark 1st digit of I.D. Number.
In row 27 mark 2nd digit of I.D. Number.
TURN CARD OVER TOP TO BOTTOM.
In row 53 mark 3rd digit of I.D. Number.
In row 54 mark the number "7."
LEAVE OTHER ROWS BLANK.
RETURN CARD TO ENVELOPE.

FINAL INSTRUCTIONS

1. Check to see that your marks on the mark sense cards are not out of their boundaries. Erase any stray markings. Return cards to the envelope.

2. Self-address the envelope containing the mark sense cards. Do not seal this envelope.

   If the envelope has become detached re-attach it to the test booklet.

3. Be sure to return the electrographic pencil with your test booklet.

APPENDIX E

Name________________________
Section No.____________________
Instructor_____________________

Personal Opinion Scale

The following is a study of what the general public thinks and feels about a number of important social personal questions. The best answer to each statement below is your personal opinion. We have tried to cover many different and opposing points of view; you may find yourself agreeing strongly with some of the statements, disagreeing just as strongly with others, and perhaps uncertain about others; whether you agree or disagree with any statement, you can be sure that many people feel the same as you do.

Indicate the extent to which you agree with each statement on these pages by marking in front of the statement one of the numbers 1 through 6 corresponding to each statement by using the following scale.

6-I agree very much	**3-I disagree a little**
5-I agree on the whole	**2-I disagree on the whole**
4-I agree a little	**1-I disagree very much**

Do not linger over any item. Your first impression is what is desired.

_____1. Obedience and respect for authority are the most important virtues children should learn.
_____2. No weakness or difficulty can hold us back if we have enough will power.
_____3. Science has its place, but there are many important things that can never possibly be understood by the human mind.
_____4. Human nature being what it is, there will always be war and conflict.
_____5. Every person should have complete faith in some supernatural power whose decision he obeys without question.
_____6. When a person has a problem or worry, it is best for him not to think about it but keep busy with more cheerful things.
_____7. A person who has bad manners, habits, and breeding can hardly expect to get along with decent people.
_____8. What the youth needs most is strict discipline, rugged determination, and the will to work and fight for family and country.
_____9. Some people are born with an urge to jump from high places.
_____10. Nowadays when so many different kinds of people move around and mix together so much, a person has to protect himself especially carefully against catching an infection or disease from them.
_____11. An insult to our honor should always be punished.
_____12. Young people sometimes get rebellious ideas, but as they grow up they ought to get over them and settle down.
_____13. It is best to use some prewar authorities in Germany to keep order and prevent chaos.
_____14. What this country needs most, more than laws and political programs, is a few courageous, tireless, devoted leaders in whom the people can put their faith.
_____15. Sex crimes, such as rape and attacks on children, deserve more than mere imprisonment; such criminals ought to be publicly whipped or worse.
_____16. People can be divided into two distinct classes; the weak and the strong.
_____17. There is hardly anything lower than a person who does not feel a great love, gratitude, and respect for his parents.
_____18. Some day it will probably be shown that astrology can explain a lot of things.
_____19. Nowadays more and more people are prying into matter that should remain personal and private.
_____20. Wars and social troubles may someday be ended by an earthquake or flood that will destroy the whole world.
_____21. Most of our social problems would be solved if we could somehow get rid of the immoral, crooked, and feebleminded people.
_____22. The wild sex life of the old Greeks and Romans was tame compared to some of the goings-on in this country, even in places where people might least expect it.
_____23. If people would talk less and work more, everybody would be better off.
_____24. Most people don't realize how much our lives are controlled by plots hatched in secret places.
_____25. Homosexuals are hardly better than criminals and ought to be severely punished.
_____26. The businessman and the manufacturer are much more important to society than the artist and professor.

_____27. No sane, normal, decent person could ever think of hurting a close friend or relative.

_____28. Familiarity breeds contempt.

_____29. Nobody ever learned anything really important except through suffering.

APPENDIX F

I.D. No.__________________
Name__________________
Section No.__________________
Instructor__________________

PART II

PERSONAL OPINION SCALE

The following is a study of what the general public thinks and feels about a number of important social and personal questions. The best answer to each statement below is your personal opinion. We have tried to cover many different and opposing points of view; you may find yourself agreeing strongly with some of the statements, disagreeing just as strongly with others, and perhaps uncertain about others; whether you agree or disagree with any statement, you can be sure that many people feel the same as you do.

Instructions

1. Place an unused mark sence card underneath the right edge of the test booklet. Place the card such that the columns of 9's, 8's, and 7's are covered by the test booklet. You may wish to lift the test booklet to check your position. The rows numbered 1 through 27 (the green side of the card) correspond to statements 1 through 27 on page F-2. The rows numbered 28-42 (the brown side of the card) correspond to statements 28 through 42 on page F-3.

2. Indicate the extent to which you agree with each statement on these pages by marking one of the numbers 1 through 6 corresponding to each statment by using the following scale.

6 I agree very much	3 I disagree a little
5 I agree on the whole	2 I disagree on the whole
4 I agree a little	1 I disagree very much

3. Use an electrographic pencil. Mark one and only one answer to each item.

4. Do not linger over any item. Your first impression is what is desired.

1. The United States and Russia have just about nothing in common.
2. The highest form of government is a democracy and the highest form of democracy is a government run by those who are most intelligent.
3. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups.
4. It is only natural that a person would have a much better acquaintance with ideas he believes in than with ideas he opposes.
5. Man on his own is a helpless and miserable creature.
6. Fundamentally, the world we live in is a pretty lonesome place.
7. Most people just don't give a "damn" for others.
8. I'd like it if I could find someone who would tell me how to solve my personal problems.
9. It is only natural for a person to be rather fearful of the future.
10. There is so much to be done and so little time to do it in.
11. Once I get wound up in a heated discussion I just can't stop.
12. In a discussion I often find it necessary to repeat myself several times to make sure I am being understood.
13. In a heated discussion I generally become so absorbed in what I am going to say that I forget to listen to what the others are saying.
14. It is better to be a dead hero than to be a live coward.
15. While I don't like to admit this even to myself, my secret ambition is to become a great man, like Einstein, or Beethoven, or Shakespeare.
16. The main thing in life is for a person to want to do something important.
17. If given the chance I would do something of great benefit to the world.
18. In the history of mankind there have probably been just a handful of really great thinkers.
19. There are a number of people I have come to hate because of the things they stand for.
20. A man who does not believe in some great cause has not really lived.
21. It is only when a person devotes himself to an ideal or cause that life becomes meaningful.
22. Of all the different philosophies which exist in this world there is probably only one which is correct.
23. A person who gets enthusiastic about too many causes is likely to be a pretty "wishy-washy" sort of person.
24. To compromise with our political opponents is dangerous because it usually leads to the betrayal of our own side.
25. When it comes to differences of opinion in religion we must be careful not to compromise with those who believe differently from the way we do.

LEAVE ROWS 26 AND 27 BLANK. TAKE CARD TO NEXT PAGE

TURN CARD OVER TOP TO BOTTOM

28. In times like these, a person must be pretty selfish if he considers primarily his own happiness.
29. The worst crime a person could commit is to attack publicly the people who believe in the same thing he does.
30. In times like these it is often necessary to be more on guard against ideas put out by people or groups in one's own camp than by those in the opposing camp.
31. A group which tolerates too much differences of opinion among its own members cannot exist for long.
32. There are two kinds of people in this world: those who are for the truth and those who are against the truth.
33. My blood boils whenever a person stubbornly refuses to admit he's wrong.
34. A person who thinks primarily of his own happiness is beneath contempt.
35. Most of the ideas which get printed nowadays aren't worth the paper they are printed on.
36. In this complicated world of ours the only way we can know what's going on is to rely on leaders or experts who can be trusted.
37. It is often desirable to reserve judgment about what's going on until one has had a chance to hear the opinions of those one respects.
38. In the long run the best way to live is to pick friends and associates whose tastes and beliefs are the same as one's own.
39. The present is all too often full of unhappiness. It is only the future that counts.
40. If a man is to accomplish his mission in life it is sometimes necessary to gamble "all or nothing at all."
41. Unfortunately, a good many people with whom I have discussed important social and moral problems don't really understand what's going on.
42. Most people just don't know what's good for them.

PLACE CARD INSIDE FRONT COVER AND RETURN COMPLETED FORM TO PERSON IN CHARGE

APPENDIX G

INTRODUCTION

The following is a study of what the general public thinks and feels about a number of important social and personal questions. The best answer to each statement below is your personal opinion. We have tried to cover many different and opposing points of view; you may find yourself agreeing strongly with some of the statements, disagreeing just as strongly with others, and perhaps uncertain about others; whether you agree or disagree with any statement, you can be sure that many people feel the same as you do.

Mark each statement according to how much you agree or disagree with it by placing the appropriate number in the box in the left-hand margin. Please mark every one. Write +1, +2, +3, or -1, -2, -3, depending on how you feel in each case.

+1: I AGREE A LITTLE
+2: I AGREE ON THE WHOLE
+3: I AGREE VERY MUCH

-1: I DISAGREE A LITTLE
-2: I DISAGREE ON THE WHOLE
-3: I DISAGREE VERY MUCH

* *

☐ 1. Most of the people I know have nothing in common with mentally ill individuals.
☐ 2. Psychologists are not the least bit like Navaho medicine men.
☐ 3. The college student who studies psychology looks at things quite differently from other students.
☐ 4. In discussing psychology, people have a way of arguing about irrelevant issues like handwriting analysis rather than pursuing the really significant questions.
☐ 5. While psychology should not be used to manipulate values and beliefs, there are situations in which propaganda techniques can be used for noble causes.
☐ 6. Even though research is necessary for the advancement of science it is abstract, unrealistic, and has few practical benefits.
☐ 7. Although the data from well-designed and well-controlled scientific experimentation are valid, the results of psychological experimentation on people are voided by the element of free will.
☐ 8. Although scientific inquiry into the nature of the world is both necessary and desirable, psychologists should stop invading individual privacy.
☐ 9. In studying psychology, people naturally pay more attention to the ideas they agree with than to the ideas they oppose.
☐ 10. Although there are many different types of people engaged in counseling who are not psychologists, it is difficult to distinguish between the amount of help they provide since they are equally inadequate.
☐ 11. Most mental illness is due to the fact that we live in a cold and overwhelming world.
☐ 12. The inevitable loneliness of man is the basis of most maladjustments.
☐ 13. The large number of mentally ill people in the world shows how badly men treat each other.
☐ 14. I would like to have a good counseling psychologist around me all of the time.
☐ 15. Only science can save us from the threats of the future.
☐ 16. Science must constantly advance every minute if it is to find the "answers" in time.

APPENDIX H

I.D. No.________________
Name____________________
Section No._______________
Instructor________________

PART I

SUBGROUP CHARACTERIZATION FORM

In everyday experience you make conscious or unconscious characterizations of groups of people. If you were to go to a meeting of German scientists you probably would say they were "very intelligent" rather than "not intelligent." If you went to a beauty contest you would be more likely to say the contestants were "attractive" rather than "unattractive."

In this questionnaire you are to indicate what you would probably say or think if you observed certain groups of people named at the top of the following pages. You are to do this by marking a position near to or far from the opposite ends of several scales like "attractive-unattractive." For example, if you feel that you would probably find a group of psychology professors very boring then on the page marked "Psychology Professor" you should indicate a position much closer to the word "boring" than to the word "interesting." If you feel that there is no relation between the group and the scale like "boring-interesting" then you should mark the middle position. Do not use the middle position, however, to avoid choosing the more likely side of the scale.

READ CAREFULLY

INSTRUCTIONS FOR USE OF IBM MARK SENSE ANSWER CARDS

1. An IBM electrographic pencil must be used.

2. Your answer must be within the bounds of the answer space and not run into the next column.

3. All changes or stray marks must be carefully erased. Do not cross out changes.

4. Mark an answer to each item. Do not mark more than one answer to an item.

5. Insert a card into a page using the slits made in the page for this purpose. Only options 1 through 7 to each item should be visible and may be used. Odd-numbered pages will use the green lettered sides of the cards. Even-numbered pages will use the brown lettered sides. The spaces marked 7 will always be on the left while those marked 1 will always be on the right.

6. A mark in the space numbered 7 will indicate that you would probably say or think the group being described would be very much characterized by the words on the left of the card inserted in the page. A mark in the space marked 1 will indicate that you would probably say or think the group being described would be very much characterized by the words on the right of the card inserted in the page. Options 6 through 2 will indicate relative positions in between these extremes.

7. Follow instructions on each page carefully.

8. Do not linger over any item. Your first impression is what is desired.

JEWISH COLLEGE PROFESSOR

7 ○ 6 ○ 5 ○ 4 ○ 3 ○ 2 ○ 1 ○

INSERT UNUSED CARD TO COVER THIS SPACE

has a pretty wife	1	wife is not pretty
has poor taste	2	has good taste
unhappy home life	3	happy home life
cleancut	4	strange
not interested in art	5	interested in art
worthless	6	valuable
deep	7	shallow
colorless	8	colorful
optimistic	9	pessimistic
hot	10	cold
interesting	11	boring
selfish	12	unselfish
wise	13	foolish
sensitive	14	insensitive
excitable	15	calm
unintelligent	16	intelligent
has no emotional problems	17	has emotional problems
persevering	18	quitting
indifferent to people	19	attentive to people
thoughtful	20	unreflective
irresponsible	21	responsible
degenerate	22	not degenerate
attractive	23	unattractive
unrespectable	24	respectable
clannish	25	eager to mix with other groups
MARK 1ST DIGIT OF ID NO. HERE	26	
MARK 2ND DIGIT OF ID NO. HERE (FIND ID NO. ON THE COVER) (Slide card if necessary to mark 8, 9, or 0.)	27	NOTE CORNER

7 ○ 6 ○ 5 ○ 4 ○ 3 ○ 2 ○ 1 ○

REMOVE CARD AND TAKE IT TO NEXT PAGE

APPENDIX I

Name____________________

STUDENT OPINION SCALE

On the following pages 35 questions are asked. Below each question you will find five alternative answers. In the spaces at the far left you are to check those answers which are acceptable to you as answers to the questions. It may be that none of these answers are acceptable to you; it may be that they all are. You are to check those which are acceptable answers to you.

In the next spaces (second column) you are to rank-order the five alternatives in terms of their acceptability. That is, put a "1" before the alternative that is most acceptable to you, a "2" before the alternative that is the next most acceptable, and so on until you have put a "5" before the alternative that is least acceptable to you.

Some of the items of this form ask very personal questions. In spite of the fact that your answers will be kept confidential and used only for research purposes you may not wish to answer these items. If you do not wish to answer a question, write this fact in the answer spaces and go on to the next question. Answer all items in some fashion; do not omit items.

I. Psychology and Science

1. To what extent is psychology a science?

____ ____a. Psychology is not a science at all.
____ ____b. A small part of the field of psychology is scientific.
____ ____c. Psychology has about equal portions of scientific material and nonscientific material.
____ ____d. Psychology is basically a science but lacks scientific rigor in some areas.
____ ____e. Psychology is as much a science as any natural science.

2. To what extent should psychology be employed to influence human behavior?

____ ____a. It should never be consciously used to influence anyone.
____ ____b. It should be used to influence only one's self.
____ ____c. It should be used to influence others in a few restricted ways.
____ ____d. It should be used to influence others as much as one considers desirable.
____ ____e. It should be used extensively to guide people in building a better life for mankind in general.

3. What should be the role of psychologists in society?

____ ____a. They should develop their field but leave its use to others.
____ ____b. They should be active in just a few areas such as education and mental health.
____ ____c. They should have the same influence in society as any other occupational group.
____ ____d. They should have an important role in deciding how society should be run.
____ ____e. They should be the most influential group in governing society.

4. To what extent should courses in psychology be required?

____ ____a. Psychology should be an elective in high school and college.
____ ____b. It should be an elective in high school but one course should be required of certain majors in college.
____ ____c. It should be an elective in high school but one course should be required of all college students.
____ ____d. A course should be required of all high school and another course should be required of all college students.
____ ____e. It should be a required course starting in grade school and continuing at least through one course in high school and one course in college.

5. To what extent should people be concerned about the uses of psychology?

____ ____a. There is nothing to be concerned about; psychology is quite harmless.
____ ____b. Psychology offers very little to worry about if people in general do not lose their heads.

____ ____c. Psychology, like any body of knowledge, provides its possessor with power which can be abused and people must be vigilant in guarding against this abuse.

____ ____d. Psychology is especially dangerous among the bodies of knowledge because it can provide considerable power over other human beings.

____ ____e. The dangers of psychology are far more grave than is generally recognized; psychology can potentially provide dangerous forms of inhumanity.

II. Man

1. To what extent is human behavior determined?

____ ____a. Man is a free agent depending only on his will to be free.

____ ____b. Man is generally free within a few limitations of the natural world.

____ ____c. Man is determined in certain respects but may exercise free will in about as many other respects.

____ ____d. Man is generally determined but can exercise free will in a few instances.

____ ____e. Man is entirely determined by laws of the natural world.

2. To what extent is man apart from the lower animals?

____ ____a. Man is basically an animal differing from the lower animals only in terms of complexity.

____ ____b. There are a few characteristics of man which set him apart from the lower animals but for the most part man is only a more complex animal.

____ ____c. The lower animal and uniquely human elements are represented in man in about equal proportion depending on the person.

____ ____d. Man in unlike the lower animals in more respects than he is like the lower animals.

____ ____e. The characteristics of man are such that analyses of animal behavior have little or no relevance to the study of human behavior.

3. To what extent should a person try to become close friends with others?

____ ____a. Be self-sufficient and don't form close ties with anyone; one doesn't get hurt that way.

____ ____b. Form close ties with only a few people who are really understanding and can be trusted.

____ ____c. Become close friends with anyone you trust. A lot of people can be trusted but a lot cannot.

____ ____d. Try to become close friends with all the people you know. Most people will be loyal friends if they know they are trusted, though a few may take advantage of such trust.

____ ____e. Let people know you trust them and want to be close friends with them; they will respond in kind.

4. Does human nature predispose man to be good or evil?

____ ____a. Man is basically good.
____ ____b. Man is predominantly good.
____ ____c. Man is neutral or equally good and evil.
____ ____d. Man is predominantly evil.
____ ____e. Man is basically evil.

5. Suppose a man is acting as a student proctor in an examination for some large, introductory course. He sees that his roommate and best friend is cheating in the exam. The roommate is a hard-working though not a brilliant student, and desperately needs a good grade in this course. The decision as to what to do about it is entirely in the proctor's hands, for no one else sees the cheating. What should the proctor do?

____ ____a. Take away his exam book and report him for cheating.
____ ____b. Prevent him from cheating further, let him finish the exam, and report him for cheating.
____ ____c. If he can be led to withdraw from the exam on some excuse, treat him as absent; otherwise report him for cheating.
____ ____d. Prevent further cheating but let him finish the exam, and not report him for cheating.
____ ____e. Act as if nothing had happened, and not report him for cheating.

III. Sex and Family

1. How much sex education should parents give their children?

____ ____a. None is needed; the child will find out all he needs to know about sex, and soon enough, from his playmates, books, etc.
____ ____b. Parents should give their children sex education prior to marriage; before that none is needed.
____ ____c. Parents should instruct their children in matters of sex before they are adolescents, so that they will have all the facts before they need them.
____ ____d. Parents should answer honestly any questions that their children ask them about sex, at any age.
____ ____e. Parents should begin at a very early age to educate their children thoroughly in matters of sex.

2. Who should make family decisions?

____ ____a. The husband has responsibility for the family, and therefore should have authority to make all family decisions.

____ ____b. The husband should make all family decisions except those involving the care and training of children, in which the wife should share.

____ ____c. The husband should make all family decisions except those about the care and training of children and those of a household nature, which should be the wife's responsibility.

____ ____d. Family decisions should be made by joint agreement of husband and wife, but the husband should make the decisions in cases where they cannot reach an agreement.

____ ____e. Family decisions should all be made by joint agreement between husband and wife.

3. Is it a good idea for a man to marry before finishing college?

____ ____a. No; college students are too young, too financially insecure, and too inexperienced to know what they want.

____ ____b. Usually it's a bad thing, because the husband cannot be the major financial support, and he is not situated in a permanent job.

____ ____c. It depends entirely on the individual case; for some couples it works out well and for others it does not.

____ ____d. Usually it's a good thing, because college couples can pool whatever resources they have, and generally are mature enough to know what they want.

____ ____e. Yes; it should be encouraged because the experience of marriage is good for the student and usually helps him settle down to work harder.

4. What are the proper standards concerning pre-marital and extra-marital sexual activity?

____ ____a. Sexual activity for either men or women before marriage, and extra-marital sexual activity after marriage, are wrong.

____ ____b. Some sexual freedom for men but not for women should be permitted before marriage, but not after marriage for either.

____ ____c. Freedom in sexual activity for both men and women should be permitted before marriage but not afterward.

____ ____d. Both sexes should be allowed sexual freedom before marriage, and some extra-marital activity is OK after marriage for men but not for women.

____ ____e. Both sexes should be allowed sexual freedom both before and after marriage, provided both partners agree on the arrangement.

5. How important is the sex relationship in building a good marriage relationship?

____ ____a. Good marriages are built entirely on grounds other than sex; the sex relationship is of minimal importance.

____ ____b. Good marriages are built primarily on grounds other than sex, but sex is an important factor.

____ ____c. Good marriages are built on several grounds, one of which is the sex relationship.

____ ____d. Good marriages are built on several grounds, but the sex relationship is the most important factor.

____ ____e. Good marriages are based almost entirely on the sex relationship; other factors will be OK if the sex relationship is OK.

IV. Religion

1. How should differences between the findings of modern science and religious beliefs be resolved?

____ ____a. Religious viewpoints that differ with the findings of modern science should be abandoned.

____ ____b. Religious viewpoints that differ with the findings of modern science should be re-examined and probably modified to be made compatible with our best current knowledge.

____ ____c. Science and religion each has its own area of applicability, and each is valid only within these limits; let the two get along together, even if in seeming contradiction.

____ ____d. When a scientific finding seems to contradict religious principles, one must find an interpretation of the finding which is compatible with known religious truths.

____ ____e. When a scientific interpretation is not in accord with known religious truths, it is therefore not valid and should be abandoned.

2. How should one regard religious beliefs other than one's own?

____ ____a. With complete appreciation and respect; the powers of the universe reveal themselves in different ways to different people.

____ ____b. With appreciation and respect, but with recognition that some religions are better than others.

____ ____c. With appreciation and respect for the most part, but recognizing that there are some points about which one can only be tolerant.

____ ____d. With tolerance but not appreciation; conflicting beliefs cannot all be right, and it is wrong not to follow the best.

____ ____e. With no tolerance for beliefs that are in error, but with tolerance for the people who mistakenly hold them.

3. To what extent should various religious sects join together, cooperate, or compete with one another?

____ ____a. All should get together, iron out their minor differences, and consolidate into a united religious force.

____ ____b. All should get together and work in cooperation, but not unite.

____ ____c. Religious groups holding similar views ought to work out differences and consolidate; but others hold totally different views, and thus could not so combine.

____ ____d. Different religious groups ought to maintain their own views, and remain separate; but they should not transgress the right of other groups to do the same.

____ ____e. If a religious group is convinced of the rightness of its creed, it must try in all peaceful ways to spread that creed to all, and thus may have to compete with other groups for members.

4. How important is being a member of a church or religious organization in leading a good life?

____ ____a. To lead a truly religious life, one must be a member of a religious group dedicated to that purpose.

____ ____b. Maybe a few exceptional men can lead virtuous lives apart, but for most men it is necessary to share their religious beliefs with others of like mind.

____ ____c. One may lead a good life alone, but belonging to a religious group makes it easier and more satisfying through sharing beliefs with others of like mind.

____ ____d. Leading a religious life is a matter of following personal conscience in every-day behavior; there is not need to belong to a formal religious group to do so.

____ ____e. Belonging to a formal religious organization can be a hindrance to leading the best religious life, by forcing a man to abandon his own convictions in favor of views which he does not really believe.

5. What should be the place of religion in every-day life?

____ ____a. Religion should pervade every area of life, and determine all life decisions.

____ ____b. Religion should provide the underlying guiding principles of life, and thus ought to help people make all kinds of decisions.

____ ____c. Religious ideals should influence many life decisions, but they do not always provide useful guides for every-day matters.

____ ____d. There are definite areas of life where religious ideals should prevail, but there are other areas where decisions must be based on reason and experience.

____ ____e. Religious views offer only ideals; real life decisions have little to do with religion, but must be judged entirely on the basis of reason and experience.

V. Public Affairs

1. What should be our policy toward Russia?

____ ____a. We should try to reach an understanding with Russia, avoiding war if at all possible, exercising great patience.

____ ____b. We should try to get along with Russia, and avoid war unless she directly attacks our own territory.
____ ____c. We should try to avoid war, but also try to keep Russia from further expansion by going to war if necessary.
____ ____d. We should be ready to go to war any time Russia starts or causes trouble anywhere in the world.
____ ____e. We should wage preventive war against Russia now, while we still have the advantage of superior weapons.

2. What should the United States do in regard to sharing atomic-energy information and resources?
____ ____a. We should keep all atomic information resources strictly to ourselves.
____ ____b. We should share nonmilitary atomic information and resources, but only with our allies.
____ ____c. We should share nonmilitary atomic information and resources with all non-Communist countries.
____ ____d. We should share nonmilitary atomic information and resources with all countries.
____ ____e. We should share nonmilitary atomic information and resources with all countries, and share military atomic information with our allies.

3. How good a president was Eisenhower?
____ ____a. An excellent president; one of our best.
____ ____b. A generally good president; most of his policies have been good.
____ ____c. A fair president; his policies have been good at times and poor at others.
____ ____d. Not a particularly good president; his policies have been bad more often than good.
____ ____e. A generally poor president; almost all his policies have been bad ones.

4. In general, how successful has Senator McCarthy been in exposing Communist operations in this country?
____ ____a. The harm he has doen is far greater than the good, if any.
____ ____b. He has probably done more harm than good.
____ ____c. The good and the harm he has done are about equally matched.
____ ____d. He has probably done more good than harm.
____ ____e. The good he has done is far greater than the harm, if any.

5. What selective service regulations does this country need?
____ ____a. No selective service; the armed forces should recruit men entirely by volunteers.
____ ____b. Limited selective service; any man who has good cause should be permanently deferred.
____ ____c. All able-bodied men should be subject to selective service, but should be able to defer their service for good cause.

____ ____d. All able-bodied men should serve their time in the armed forces when their number is called; there should be no deferments, temporary or permanent, for anything but physical health.

____ ____e. All able-bodied men in the country should serve their time in the armed forces, and then remain permanently on call for service.

VI. Race Relationships

1. In what ways, if any, can Federal laws aid in handling problems of Negro-White conflict?

____ ____a. Make it illegal for persons to discriminate against Negroes in any manner.

____ ____b. Make it illegal for persons to discriminate against Negroes in the areas of housing, employment, and education, but otherwise leave things alone.

____ ____c. Make a few general laws, but mostly we should leave the problem up to individuals.

____ ____d. Make no laws concerning discrimination, but leave it to the people to work out their own ways of handling this conflict.

____ ____e. Accept the fact that there will always be Negro-White conflict, and make realistic laws calling for segregation.

2. How should the problem of assigning Negroes and Whites to schools be handled?

____ ____a. No segregation of schools; Negroes and Whites should attend the same schools and be treated equally by teachers and by other children.

____ ____b. Separate schools for those who want them, whether Negroes or Whites, and unsegregated state-supported schools for all others.

____ ____c. Negroes and Whites should attend the same schools and be treated equally for academic purposes, but separated for social functions.

____ ____d. Negroes and Whites should attend the same school buildings, but should have separate classrooms and other facilities within those buildings.

____ ____e. Negroes and Whites should have separate, state-supported schools, but the state should rigorously enforce real equality in facilities.

3. What about laws concerning marriages between Negroes and Whites?

____ ____a. There should be strict laws against Negro-White marriages in all states.

____ ____b. Laws against-Negro-White marriages are good for southern states, but are not needed elsewhere.

____ ____c. No laws against Negro-White marriages, but both parties to such marriages should be condemned.

____ ____d. Laws against Negro-White marriages are wrong, but such marriages should be discouraged, since the children of them will suffer.
____ ____e. Laws against Negro-White marriages are wrong; every man has a right to marry anyone he pleases without social disapproval.

4. Who is to blame for continuing conflict between Negroes and Whites?
____ ____a. Negroes are entirely to blame; their "slave complex" has made them very hostile to Whites.
____ ____b. Negroes are mostly to blame, but attempts by Whites to "keep them in their place" are also partly to blame.
____ ____c. Negroes and Whites are equally to blame for the continuing conflict.
____ ____d. Whites are mostly to blame, but Negroes have invited this treatment, by their behavior.
____ ____e. Whites are entirely to blame; they have imposed discrimination and justify any device to keep Negroes from getting ahead.

5. Whites should be willing to accept Negroes into relationships of what degrees of familiarity?
____ ____a. To close kinship by marriage.
____ ____b. To close personal friendship.
____ ____c. To the same street as neighbors.
____ ____d. To membership in the same profession.
____ ____e. To admission in the same movie theatres.

III. CHRONOLOGY AND DESCRIPTION OF SAMPLES USED

1. SPRING SEMESTER, 1959-60

The Peer Rating form used by Norman had been filled out by teaching fellows in the introductory psychology course during the fall semester. Five scores were obtained and correlated with ratings by students of the teachers' ability to teach.

In the spring semester the same teaching fellows filled out Norman's Descriptive Adjective Inventory (self-report) and the Cattell 16 PF Questionnaire. Scores from these two instruments, plus the Peer Ratings, were used in a multiple regression equation to predict mean ratings of teachers by their students.

A study was also carried out in The University of Michigan School of Education using the Peer Rating form and the Guilford-Zimmerman Temperament Survey to correlate with ratings of student teachers by their supervisors.

2. FALL SEMESTER, 1960-61

Eight forms of the Criteria Test were administered as pre-tests and post-tests to the students in introductory psychology. Item analyses of the tests were carried out, and the number of forms reduced to four.

The Peer Rating form, the Descriptive Adjective Inventory, and the Cattell 16 PF were administered to the teaching fellows in the course. Scores on these instruments were correlated with students' ratings of their teachers' overall teaching ability. Peer Group ratings and student ratings of teachers of introductory psychology at Ohio State University were also obtained, and results compared with Michigan findings.

An extensive student rating form of 145 items, taken from a number of forms that had previously been used here and elsewhere was filled out by a sample of 150 men and 150 women students. Item scores were factor analyzed and a 46-item scale prepared which contained the items with highest factor loadings.

3. SPRING SEMESTER, 1960-61

The Criteria Test was used as a pre-test and as a post-test for the students in the introductory course. An additional group of 100 students at Eastern Michigan University took it as a post-test. Classroom observations were made by observers trained in the Bales-Mann technique for recording group interactions and by other observers using a rating scale developed in this project. Students were tested for Need Achievement, Need Affiliation, and Need Power by means of the TAT, and took the Cattell 16 PF, an adaptation of the Brown-Holtzman Survey of Study Habits and Attitudes, the Alpert-Haber Anxiety Scale, A Vocational Orientation Questionnaire, the F scale, and a number of other attitude scales. At the end of the semester they rated their instructors. All these measures were studied in relation to each other and to three outcome measures: course grade, Criteria Test score, and score on a 15 item test of knowledge of content.

4. FALL AND SPRING SEMESTERS, 1961-62

The main activity in the project this year was the analysis of the data collected in the spring semester of 1961. The Criteria Test was given in the fall semester to the students in Psychology 100 and 101, and to Psychology 100 plus one section of 101 in the spring. On the basis of available data the number of forms of the test was reduced to two. Each form now has 60 items, with 10 items allocated to each of 6 objectives.

A sample of senior concentrates in various departments of the University's colleges of Literature, Science, and the Arts, and of Engineering were queried as to their concepts of the characteristics of ideal teachers. Their replies were related to the student rating data we had collected.

5. FALL SEMESTER, 1962-63

At the beginning of the semester the project staff participated in a two-day conference for the teachers in introductory psychology. Our part in the conference was to describe our research and explain our plans for the 1962-63 academic year.

The major activity for the first semester was assessing the implications of the data we had collected thus far with a view toward determining how students should be placed with teachers so as to maximize their probability of good performance. Three main categories of teacher behaviors were identified—Achievement Cues, Skill, and Culture—and student traits that interacted with them found. On the basis of a median split in each category eight "types" of teachers and of students could be labeled. It was thus possible to say, for any given student-teacher pair, how good a match there was with respect to the trait categories.

The University was moving into a system of pre-registration, so that it was possible for us to identify ahead of time about 520 students who would be taking Psychology 100 or 101 in the spring. These students were pre-tested with the TAT, the "Sophistication" (N) scale of the 16 PF, the study habits inventory, the Alpert-Haber Anxiety scale, and two attitude questionnaires. Scores from the freshman orientation battery (SAT and OAIS) were available for students who had entered the University as freshmen. On the basis of these pretest data each student could be classified as one of the eight types.

Student ratings of teachers, made at the end of the fall term, provided the measures of the teachers on the three categories, and they, likewise, could be classified into one of eight types. The teachers also took the Descriptive Adjective Inventory, but not the Peer Rating form, as we had intended originally.

Both new forms (X and Y) of the Criteria Test were given as a part of the final examinations in Psychology 100 and 101. Intercorrelations of subtests within and between forms were studied and a factor analysis carried out.

6. SPRING SEMESTER, 1962-63

In Psychology 101, at least two sections of the course had been scheduled each hour, so that students could be shifted from one to the other without disrupting their own schedules. As much as was possible students were shifted so as to obtain the best matches between them and their teachers. Outcome measures consisted of course grades, the Criteria Test, a 25-item objective test of knowledge of content, and an essay test.*

In Psychology 100 enrollment is smaller than in 101, so it was not possible to shift students from one section to another, and the matching of student to instructor was not manipulated. Outcome measures were course grades and the Criteria Test.

A factor analysis of student ratings of instructors in the introductory course in Economics was also carried out this semester.

---

*The objective test and the essay test were prepared by Karl Zinn in connection with his doctoral dissertation. Data relating to them will appear in his report.

Description of Samples Used

Sample	N (Approximate)	Measures
Teaching fellows in Introductory Psych., Fall, 1959-60	15	1. Ratings by other Teaching fellows 2. Ratings by students
Teaching fellows in Introductory Psych., Spring, 1960	15	1. Ratings by other Teaching fellows 2. Descriptive Adjective Inventory (self-report) 3. Cattell 16 PF
Student teaching seminar, U. of M. School of Education, Spring, 1960	50	1. Ratings by peers 2. Evaluations by seminar leader 3. Evaluations by practice teaching supervisors. 4. Guilford-Zimmerman Temperament Survey 5. Ratings of the deportment and classroom behavior of the student teachers' pupils
Students in summer school, 1960 classes in Introductory Psych. at U. of M. and N. C. State College	52	1. Criteria Test 2. Adjective Check List, adapted from Sommer and Killian
Students in summer school, 1960 classes in Introductory Psych. at MSU	33	1. Criteria Test
Students in Introductory Psych. at U. of M., Fall, 1960-61	1500	1. Criteria Test: 8 forms distributed; administered at beginning and end of course
Students in Introductory Psych. Subject Pool at U. of M., Fall, 1960-61	300	1. Student Rating Form
Students in Introductory Psych. at OSU, Fall, 1960-61	925	1. Student ratings of teachers

Sample	N (Approximate)	Measures
Teaching fellows in Introductory Psych. at OSU, Fall, 1960-61	11	1. Peer Group ratings by other teaching fellows 2. Descriptive Adjective Inventory (self-report) 3. Cattell 16 PF
Students in Introductory Psych. Subject Pool at U. of M., Fall, 1960-61	200	1. Adjective Check List derived from Sommer and Killian, for "person" and "Negro" 2. Syllogisms (adapted from Thistlethwaite) 3. Dogmatism Scale (adapted from Rokeach)
Teaching fellows in Introductory Psych., Fall, 1960-61	15	1. Ratings by other teaching fellows 2. Cattell 16 PF 3. Descriptive Adjective Inventory (self-report)
Students in Introductory Psych., Spring, 1960-61	750	1. Criteria Tests: pre-test and post-test 2. TAT: n Ach, n AFF, n Power 3. Cattell 16 PF 4. Study Habits Inventory, adapted from Brown-Holtzman SSHA 5. Alpert-Haber Anxiety Scale 6. California F Scale 7. Vocational Orientation questionnaire 8. Questionnaire on course and college goals 9. Measures of attitude toward minority groups 10. Classes observed 11. Ratings of instructors 12. 15-item test of knowledge of Psych. 13. Course grade
Students in Introductory Psych., at EMU, Spring, 1960-61	100	1. Criteria Test
Students in Introductory Psych., Fall, 1961-62	1150	1. Criteria Test 2. Ratings of instructors

Sample	N (Approximate)	Measures
Students in Psych. 100, Spring, 1961-62	189	1. Criteria Test
Students in one section of Psych. 101, Spring, 1961-62	130	1. Criteria Test
Teaching fellows in Introductory Psych., Fall, 1962-63	19	1. Descriptive Adjective Inventory
Students in Introductory Psych., Fall, 1962-63	1450	1. Criteria Test Forms X and Y 2. Ratings of instructors
Students in Introductory Psych., Spring, 1962-63	825	1. TAT: n Ach, n Aff, n Power 2. Cattell N 3. Alpert-Haber Anxiety Scale 4. Study Habits inventory 5. Two attitude questionnaires 6. Criteria Test, Form X 7. Objective test of knowledge of content (Psych. 101 only) 8. Essay test (Psych. 101 only) 9. Course grade
Students in introductory Economics, Spring, 1962-63	300	1. Ratings of instructors

IV-A. THE DIMENSIONS OF STUDENT EVALUATIONS OF TEACHING

Robert L. Isaacson, Wilbert J. McKeachie, John E. Milholland, Yi G. Lin,
Margaret Hofeller, James W. Baerwaldt, and Karl L. Zinn

If we are to select or train teachers effectively we need to know what teacher behaviors make differences in achievement of educational goals by their students. But before we can identify which behaviors are crucial, we need to know the dimensions on which teachers show variation. This paper deals with the problem of the identification of the dimensions of teacher behavior as perceived by their students.

During the last 40 years many scales have been devised for rating characteristics of teaching (see Ryans, 1963 for a recent review). These scales include hundreds of different items, many of which are closely related. Fortunately several of the scales have been factor analyzed. Three studies report factor analyses of the 10 items comprising the Purdue Rating Scale. The first, a study by Smalzreid and Remmers (1943) used student ratings of 40 high school teachers. The second analyzed ratings of 65 instructors at Purdue University (Creager, 1950). The third used 11 instructors at the University of Pittsburg (Bendig, 1954). All three studies found two major factors. One of these contained items such as "Sympathetic Attitude Toward Students" and "Fairness in Grading." The other contained items like "Self-Reliance and Confidence," "Presentation of Subject Matter," and "Interest in Subject Matter." In addition to these factors, Bendig also extracted a general factor with high loadings from almost all items.

In an earlier study Bendig (1953) carried out an inverse factor analysis of the Miami University faculty rating scale and found three factors.

The first dimension was characterized at one end by instructors whose classes were interesting, and smoothly controlled, and who had a good sense of humor. The second factor was characterized by instructors whose course material was loosely organized but who had very friendly attitudes toward students and were fair in their assignments. The third factor was characterized at one extreme by instructors who were rated as appreciative of student efforts and attractive in appearance but whose examinations are sometimes unfair.

Gibb (1955) developed a teacher behavior rating instrument based on social psychological analyses of leader behavior. Factor analysis of his nine scales revealed four identifiable factors. Factor I he called "Friendly Democratic Behaviors." Factor II was labeled "Communication Behavior" and characterized the teacher who facilitated exchanges of information in the group. Factor III was labeled "Organization Behavior" and characterized the instructor who was systematic and businesslike. Factor IV was "Academic Emphasis." Gibb's Factor I appears to be similar to the first factor in the studies of the Purdue scale, but the second factor on the analyses of the Purdue scale is less clearly matched in Gibb's analysis.

From these analyses it would appear that one dimension on which college teachers differ has to do with friendly, sympathetic relationships with students. Other dimensions have not been as clearly established. Only Gibb's scale contained enough items to permit very extensive sampling of the domain of teacher behavior. The present study was designed to replicate and extend Gibb's work by including Gibb's items with other items in ratings of new samples of college teachers.

PROCEDURE

In the Fall of 1960 a sample of about 150 men and 150 women students was drawn in various sections of the introductory psychology course at The University of Michigan. Each student rated his teacher on 145 items culled from many different student rating instruments. Some of these items came from the Teacher Behavior Description inventory developed by Gibb, others from student evaluation used by the psychology department at Ohio State University and by the University of Minnesota (Student Report on Classroom Teaching, Form B), as well as from student evaluation devices used formerly by the Department of Psychology, The University of Michigan. The item responses of this sample were divided into groups of about 50 items and factor analyzed separately for each sex. Seven factors were obtained in all of these analyses and the two or three highest loading items from each factor for each sex were selected and another factor analysis performed. A total of 50 items were selected. This subsequent factor analysis produced eight factors for both the men and the women. Items which loaded on several factors were eliminated and a comparison of the factors arising in the separate analyses of men and women was made. Thirty-four items which loaded highly on one of the men's and women's factors were incorporated in the new evaluation form presented in Table 1. Additional items included were two concerned with the students' evaluations of instructor and course (items 1 and 2, Table 1), three items from an earlier student evaluation instrument not included in the factor analyses (items 4, 29, 36), four items which were included to determine the relation between the scores obtained from a peer-nomination procedure carried out by the teachers (see Isaacson, McKeachie, and Milholland, 1963) and student perceptions of the same variables (items 18, 32, 35, 36), two items to determine the students' changes related to more applied psychological matters (items 25 and 46), and one new item designed to ascertain a measure of student interaction in the classroom (item 10).

TABLE 1

STEMS OF THE 46 ITEMS WHICH COMPRISED THE STUDENT EVALUATION INSTRUMENT

1. How would you rate your instructor in general (all-around) teaching ability?

   1. an outstanding and stimulating instructor
   2. a very good instructor
   3. a good instructor
   4. an adequate, but not stimulating instructor
   5. a poor and inadequate instructor

2. How would you rate the over-all value of this course?

   1. superior
   2. very good
   3. good
   4. fair
   5. poor

3. The students in the class were friendly.

4. Class time was well spent.

5. The instructor was skillful in observing student reactions.

6. He maintained definite standards of student performance.

7. In his class, I felt free to express my opinions.

8. The instructor assigned very difficult reading.

9. He encourages class members to work as a team.

10. Students argued with one another or with the instructor, not necessarily with hostility.

11. The instructor ruled the class with an iron hand.

12. The student was able to get personal help in this course.

13. The instructor invited criticism of his acts.

14. He anticipated student difficulties before they arose.

15. He decided in detail what should be done and how it should be done.

TABLE 1 (Continued)

16. He listened attentively to what class members had to say.

17. He imposed his own goals on the students.

18. He exhibited a high degree of cultural attainment.

19. He put his material across in an interesting way.

20. He let the class set its own goals.

21. He was friendly.

22. He followed an outline closely.

23. He tried to increase the interest of class members in his subject.

24. He asked for more than students could get done.

25. His efforts improved the ability of the students to deal effectively with their personal problems.

26. He announced exams in advance.

27. He told students when they had done a particularly good job.

28. He was sensitive to a student's desire to ask questions.

29. He continually emphasized grades.

30. He made it clear how each topic fit into the course.

31. He assigned a great amount of reading.

32. He was permissive and flexible.

33. He explained the reasons for his criticisms.

34. He had everything going according to schedule.

35. He was talkative.

36. The students frequently volunteered their own opinions.

37. He explained why he did things.

TABLE 1 (Concluded)

38. He complimented a student on his work in front of others.

39. He planned the activities of each class period in detail.

40. He stimulated the intellectual curiosity of his students.

41. He changed his approach to meet new situations.

42. By the way he acted, he made the students feel afraid of him.

43. He was aware of it when students failed to keep up with him in class.

44. He criticized poor work.

45. He explained clearly and his explanations were to the point.

46. The instructor's efforts improved the ability of the students to understand deviant individuals.

The resulting evaluation instrument consisted of 46 rating items and one "additional comments" question. The first two questions, asking for the students' overall evaluation of the teacher and the course, used a multiple-choice format, and the balance of the items used a five-point rating scale. On all but the first two questions the students were asked to respond in terms of the frequency of the occurrence of certain behavioral acts in the classroom and not in terms of whether or not they considered the behavior as desirable or undesirable. The scale went from "almost always occurred" to "almost never occurred."

This report presents factor analysis of results from the use of this new evaluation instrument in the spring and fall semesters of 1961 in all sections of the introductory psychology course. Factors in the several analyses were related to each other by means of Kaiser's (1960) factor similarity technique. The factor analysis program provided a principal axes solution with Varimax rotation, and the computer was instructed to obtain as many factors as necessary to account for 95% of the response variance. Data from men and women were analyzed separately both semesters and factor similarity coefficients were obtained among the four analyses. Conceptually the factor similarity coefficient is analogous to a correlation coefficient since it is based on the cosine of the angle between two factor vectors. Because sampling distributions have not been developed for the factor similarity coefficient, it is impossible to make a statement about its statistical significance. Nonetheless, it seems to be safe to assume that coefficient values close to ±1 are indicative of very strong relationships.

The group of teachers evaluated in the Fall of 1961 were entirely different from those evaluated in the spring semester, 1961. The students were assured their reactions would be treated so as to preserve their anonymity. The form was administered during a regular classroom session. All students had ample time to complete the form. Data from 240 men students and 329 women students were obtained in the spring term and from 297 men and 392 women in the fall.

RESULTS AND DISCUSSION

We obtained similarity coefficients for every factor from one factor analysis with each factor in every other factor analysis. Table 2 summarizes these results. The first column of the table lists the factor analysis which was made the basis for the comparisons. Each cell contains a factor similarity coefficient and a number which specifies the factor from another analysis which generated the similarity coefficient. In other words, each cell entry tells which factor of one analysis is "most similar" to a given factor in the first column of the table. Thus, the first entry in column FS indicates that male spring factor MS 1 had a higher similarity coefficient (.94) with female spring factor FS 1 than any other of the female spring factors. It will be noted that the "Factor 1's" of each study all

TABLE 2

SIMILARITY ANALY S FOR THE FOUR 1961 FACTOR ANALYSES: MEN, SPRING (MS), WOMEN, S RING (FS), MEN, FALL (MF), AND WOMEN, FALL (FF)

Factor Analysis and Factors For Which Similarity Coefficients Were Calculated		Similarities With Factors in Other Analyses							
		MS		FS		MF		FF	
		Coeffici	Factors	Coefficient	Factors	Coefficient	Factors	Coefficient	Factors
MS	1. Skill			.94	1	.92	1	.97	1
	2. Overload			.98	2	.95	2	.98	2
	3. Structure			.97	3	.93	3	.96	3
	4. Evaluation			.88	4	.95	4	.91	6
	5. Interaction			.54	5	.87	5	.72	5
	6. Rapport			.82	6	.88	6	.56	5
	7.			.58	8	.58	7	.75	7
	8.			.62	7	.63	9	.69	8
	9.			.62	9	.53	7	.90	4
FS	1. Skill	.94	1			.98	1	.99	1
	2. Overload	.98	2			.98	2	.98	2
	3. Structure	.97	3			.98	3	.97	3
	4. Evaluation	.87	4			.78	4	.64	6
	5. Interaction	.54	5			.81	5	.78	5
	6. Rapport	.82	6			.94	6	.53	7
	7.	.62	8			.82	7	.66	8
	8.	.66	5			.62	8	.62	9
	9.	.62	9			.54	9	.64	7
MF	1. Skill	.92	1	.98	1			.98	1
	2. Overload	.95	2	.98	2			.98	2
	3. Structure	.93	3	.98	3			.98	3
	4. Evaluation	.95	4	.78	4			.95	4
	5. Interaction	.87	5	.81	5			.76	5
	6. Rapport	.88	6	.94	6			.70	9
	7.	.58	7	.82	7			.70	4
	8.	.60	8	.62	8			.76	8
	9.	.63	8	.56	5			.86	7
FF	1. Skill	.97	1	.99	1	.98	1		
	2. Overload	.98	2	.98	2	.98	2		
	3. Structure	.96	3	.97	3	.98	3		
	4. Evaluation	.90	9	.53	9	.95	4		
	5. Interaction	.72	5	.78	5	.76	5		
	6. Rapport	.91	4	.64	4	.70	9		
	7.	.75	7	.64	9	.86	9		
	8.	.69	8	.66	7	.76	8		
	9.	.58	8	.62	8	.52	8		

have their highest similarity coefficients with "Factor 1's" from the other factor analyses. Corresponding relations hold for Factors 2, 3, and 5. Factor 4 exhibits four exceptions in the 12 comparisons, Factor 6 has six exceptions. The remaining three factors are relatively inconsistent. The high degree of consistency and the high factor similarity coefficients across the four factor analyses suggests stability among the first six factors.

Since Kaiser's method for computing factor similarities is not well known, the complete matrix of factor similarity scores for one comparison (men, Spring, 1961 with men, Fall, 1961) is presented in Table 3. It is hoped that this will provide the reader with an idea of the type of data used to compile the results presented in Table 2. Comparisons of this kind illustrate a case of maximum difference between groups since in each population students and teachers are completely different. The factors which have been labeled tend to show higher similarity coefficients with their labeled counterparts than do the unlabeled factors. Furthermore, the less stable unlabeled factors tend to be more diluted than do the more stable labeled factors. In Table 3, for example, Factors 7, 8, and 9 have a number of medium size coefficients with other factors, more than any of the labeled factors.

The factor similarity analysis suggests that six factors can be regarded as evidencing stability over sexes and evaluation periods, student groups, and teacher groups. The items loading on them are shown in Table 4.

Factor 1 in all of the factor analyses accounted for the greatest amount of variance of the ratings. Seventeen of the items in the evaluation form were found to be highly loaded on this factor for every group. We have thought the items describing Factor 1 to be a set related to general teaching skills. They describe a teaching pattern in which material is put across in an interesting way (item 19), the intellectual curiosity of the students is stimulated (item 40), things are explained clearly (item 45), and the teacher is skillful in observing student reactions (item 5). Endorsement of these teaching qualities goes along with favorable evaluation of "all around teaching ability" (item 1) and with item 2, which asks the student to indicate the all-around value of the course.

The number and types of items which cluster together to form Factor 1, indicate that this factor represents what is often called a "halo effect" in student evaluations. In fact we found that some shorter student evaluation forms used in our department were comprised solely of items on this factor. It seems to correspond with the "Professional Impression" factor found in analyses of the Purdue scale and to Gibb's "Communication" factor. We have labeled Factor 1 as the "Skill Factor," representing both the types of items from which it stems and the loading of item 1 (Ability) on it.

TABLE 3

SIMILARITY COEFFICIENTS* BETWEEN FACTORS OBTAINED FROM MEN STUDENTS IN SPRING AND FALL OF 1961

Spring Factors	Fall Factors 1 Skill	 2 Overload	 3 Structure	 4 Feedback	 5 Interaction	 6 Rapport	 7	 8	 9
1. Skill	.92	.02	.04	.09	.01	-.12	.28	-.07	.20
2. Overload	-.03	.95	.06	.05	.00	-.02	.00	-.28	-.09
3. Structure	.08	-.04	.93	.04	-.04	.05	-.34	.06	-.04
4. Feedback	.04	-.02	-.10	.95	.10	-.17	-.17	.14	-.04
5. Interaction	.03	-.06	-.04	-.05	.87	.16	-.25	-.30	.22
6. Rapport	.01	.11	.04	.17	.03	.88	.30	.30	.06
7.	-.11	-.20	.22	.14	.17	-.02	.58	-.43	-.57
8.	.22	.13	-.10	-.19	.34	-.11	-.12	.60	.63
9.	-.28	.13	.24	-.03	.29	-.38	.52	.41	.42

*The similarity coefficients below the diagonal appeared when the spring factors were used as the basis for the comparisons, those above the diagonal when the fall factors were the basis.

TABLE 4

LOADINGS OF .30 OR MORE OF STUDENT RATING FORM ITEMS ON THE FIVE STABLE FACTORS APPEARING IN ANALYSES CARRIED OUT WITH MEN (MS) AND WOMEN (FS) STUDENTS IN INTRODUCTORY PSYCHOLOGY IN THE SPRING SEMESTER, 1961 AND IN THE FALL SEMESTER, 1961 (MF AND FF)

Item	Factor 1 (Skill)				Factor 2 (Overload)				Factor 3 (Structure)				Factor 4 (Feedback)				Factor 5 (Interaction)				Factor 6 (Rapport)			
	MS	FS	MF	FF	MS	FS	MF	FF	MS	FS	MF	FF	MS	FS	MF	FF	MS	FS	MF	FF	MS	FS	MF	FF
1	82	86	83	86																				
2	34	59	67	67																				
3																	38			36			35	
4	63	67	68	68					37															
5	57	64	59	78																				
6	37	44	39	49												33								
7		51															56			58			43	
8					68	70	68	65																
9	37			44			51		-30															
10																	38	49	49					
11																								
12	31	35		32									32										50	43
13	37			32									46									38	37	62
14	57	55	56	64																				32
15	46			34					42	60	40	43												
16		33		43																33	41	46	56	
17																								
18	57	57	44	52																				
19	86	85	81	88																				
20				30					35															
21	39	35		39																30	35	38	50	
22									64	69	70	70												
23	64	49	49	72																		38	36	
24					54	68	66	60													-30			
25	48	53	51	60										38										
26																					33	35		
27													43	37	38	32						45		
28															32	31					36	40		
29																								
30	56	47	57	67					32	33														
31					70	67	70	71																
32																					51	30	40	31
33	56	46	38	56																	42	35	47	
34	32		32	48					61	70	58	50												
35	48	30																			38	40		
36		37		31													30	49	43	50	47	35	36	
37	56	44	46	66																		51	31	
38	35			32									42	49	54	51								
39	36								58	71	68	69												
40	76	81	74	85																				
41	69	67	56	74																				
42	-30			-34																-36	-56	-30		
43		35	35	38													42							
44															36	33								
45	78	71	60	81																				
46	53	42	47	60																				

Our second stable factor (Factor 2) clearly deals with the amount and difficulty of the work the instructor expects of his students. The three items consistently appearing on this factor are:

8. Assigned very difficult reading.
24. Asked for more than students could get done.
31. Assigned a great deal of reading.

We call this the "Overload" factor. It appears to be similar to Gibb's "Production" scale.

A third stable factor (Factor 3) deals with the instructor's organization and planning of the course. Four items loaded well on this factor:

15. He decided in detail what should be done and how it should be done.
22. Followed an outline closely.
34. Had everything going according to schedule.
39. Planned the activities of each class period in detail.

We call this the "Structure" factor. It appears to be similar to Gibb's "Organization" factor.

The fourth stable factor (Factor 4) contains items which suggest that the teacher voices concern over the quality of the students' work. Two items always associated with this factor were:

27. He told students when they had done a particularily good job.
38. He complimented students in front of others.

One other item appeared to be associated with this factor:

44. He criticized poor work.

These, along with the type of other items appearing on this factor suggest that it is related to the instructor's evaluative comments about classroom performance and it was labeled the "Feedback" factor.*

The fifth stable factor (Factor 5) found in the ratings was one which we think of as a "Group Interaction" dimension. It is identified by items:

---

*The reader of our earlier article (Isaacson, McKeachie, and Milholland, 1963) will note that these items were then included in the "Student Rapport" factor, a rather diffuse collection of items. The new grouping resulting from comparisons of the four factor analyses, seems more cohesive.

3. The students in the class were friendly.
7. In his class, I felt free to express my opinions.
10. Students argue with one another or with the instructor, not necessarily with hostility.
36. The students frequently volunteered their own opinions.

A Student-Teacher rapport factor (Factor 6) was also found rather consistently, appearing in three of our four analyses. This factor was marked by the regular appearance of four items:

16. He (the teacher) listened attentively to what class member had to say.
21. He was friendly.
32. He was permissive and flexible.
33. He explained the reasons for his criticism.

This factor may be similar to Gibb's Factor 1 "Friendly Democratic Behavior" which included items on initiation membership, and integration. These items did not survive our original item selection procedures.

On each of the analyses three less stable factors appeared. Each of these "unstable" factors in a given analysis was compared with those appearing in the other analyses. None from one analysis table corresponded with any factor appearing in another. It is rather surprising that this procedure revealed no factors common only to men or to women and no factor common only to the semester under analysis.

Our results have generally confirmed the factors identified in studies prior to ours, but previous studies were limited to fewer items than the original pool with which we started. The fact that 34 of our final 46 items were the residue from the 145 we used at the outset (which themselves had seen previous service) gives us some confidence that we covered the domain adequately.

SUMMARY

Two groups of students in introductory psychology rated their teachers on a 46-item questionnaire, derived largely from factor analyses of 145 items that had been used in previous studies elsewhere. The results were factor analyzed separately by sex and semester, and factor similarities obtained by Kaiser's method. Six factors appeared which were consistent over time, student groups and teacher groups. They were labeled Skill, Overload, Structure, Feedback, Group Interaction, and Student-Teacher Rapport.

References

Bendig, A. W. An inverted factor analysis study of student-rated introductory psychology instructors. J. exp. Educ., 1953, 21, 333-336.

Bendig, A. W. A factor analysis of student ratings of psychology instructors on the Purdue scale. J. educ. Psychol., 1954, 45, 385-393.

Creager, J. A. A multiple factor analysis of the Purdue Rating Scale for Instructors. M.S. Thesis, Purdue, 1950.

Gibb, C. A. Classroom behavior of the college teacher. Educ. Psychol. Meas., 1955, 15, 254-263.

Isaacson, R. L., McKeachie, W. J., and Milholland, J. E. Correlation of teacher personality variables and student ratings. J. educ. Psychol., 1963, 54, 110-117.

Kaiser, H. F. Relating factors between studies based upon different individuals. Dittoed preliminary draft, Bureau of Educational Research, University of Illinois, July, 1960.

Ryans, D. G. Assessment of teacher behavior and instruction. Rev. educ. Res., 1963, 33, 415-441.

Smalzreid, N. T. and Remmers, H. H. A factor analysis of the Purdue Rating Scale for Instructors. J. educ. Psychol., 1943, 34, 363-367.

Journal of Educational Psychology
1963, Vol. 54, No. 2, 110-117

CORRELATION OF TEACHER PERSONALITY VARIABLES AND STUDENT RATINGS[1]

ROBERT L. ISAACSON, WILBERT J. McKEACHIE, AND JOHN E. MILHOLLAND
University of Michigan

A peer group nomination procedure, a descriptive adjective inventory, and the IPAT 16 PF Questionnaire were administered to 2 groups of teaching fellows in the introductory psychology course at the University of Michigan. These personality scores were related to the teaching fellows' teaching effectiveness ratings made by their students and to the factor scores obtained from factor analyses of the separate items of the student rating instrument. The teacher personality variable most consistently correlated with good ratings by students was the peer group evaluation of the teaching fellows' "general cultural attainment." The relations of the personality scales to the other dimensions of student ratings are also presented.

Stable correlates of teaching success have been hard to find. Studies of teacher personality characteristics related to successful teaching have often yielded conclusions which needed stringent qualifications. For example, Della Piana and Gage (1955) found that students' evaluations of teachers were related to the teachers' scores on the Minnesota Teacher Attitude Inventory (MTAI) *if* the classes were oriented generally toward the affective merits of the teachers. The student ratings of classes in which the predominant orientation was not in this direction did not correlate with the teachers' MTAI scores. These findings suggest an explanation for the lack of correlation between the MTAI teacher scores and student evaluations when the students involved were all enrolled in a university high school and presumably more oriented toward cognitive rather than affective teacher qualities (Gage & Suci, 1951).

Gowan (1955) stated that persons making high scores on the *K* scale of the Minnesota Multiphasic Personality Inventory (MMPI) show a tendency to enter into empathic relationships and to make good teachers and counselors. In addition, teachers identified as well-adjusted were thought to be the better teachers. Symonds and Dudek (1956), evaluating personality with the use of the Rorschach, reported that superior personal organization, good adjustment and reasoning, capacity to relate to others, and less aggressive responses correlated with successful teaching. In his recent book Ryans (1960) contrasted traits characteristic of teachers high and low on three scales of desirable teaching characteristics. Among the traits distinguishing the good teacher were emotional stability, friendliness, cooperativeness, agreeableness, restraint, and objectivity.

Practically the only agreement among researchers on teacher characteristics is that all of the traits of effective teachers are valued in our culture, although we should pause to note the suggestion that neurotic individuals *can* be effective, given appropriate circumstances of work (see the discussion of Symonds, 1950, pp. 692-693).

It may be that only certain traits from clusters of ones with high cultural eval-

---

[1] This research was supported in part by Office of Education Research Contract O. E. No. SAE-8451, to W. J. McKeachie, J. E. Milholland, and R. L. Isaacson.

uation will predict success in college instruction. At the time the study reported in this paper was undertaken, a technique of assessment of five culturally valued traits was being investigated by Norman (1961). This technique uses peer group nominations on 20 individual bipolar scales which are combined to yield scores on five personality factors. It stemmed from the work of Cattell (1947) and was further developed by Tupes (1957) and Tupes and Christal (1958, 1961). The factors defined by the peer nomination procedure are labeled:

1. Surgency (assertive, frank, energetic, talkative, etc.)
2. Agreeableness (cooperative, attentive, mild-mannered, not jealous, etc.)
3. Dependability (responsible, conscientious, orderly, etc.)
4. Emotional Stability (calm, poised, in control of emotions, etc.)
5. Culture (artistic, polished, imaginative, effectively intelligent, etc.)

Because all five factors describe traits relevant to teaching we believed that the peer nominations might be useful in identifying specific personality traits correlated with effective college teaching.

PROCEDURE

Teacher Sample

To obtain a sample of college-level instructors who were relatively homogeneous in age, general abilities, teaching experience, and instructional media, we secured the participation of the teaching fellows in the introductory sychology course at the University of Michian. In 2 successive years two different groups f young men and women who had not taught t the University of Michigan before became vailable.

In their initial instructional experience the aching fellows led discussion sections of a rge lecture course. In the second semester ach was entirely responsible for his own class nd both lectured and led discussion sections. We could therefore study the personality variables related to success under two types of course organization, although the second semester experience was necessarily confounded with experience derived from the first term.

All of the teaching fellows in the introductory psychology course during the 1959-60 and the 1960-61 academic years cooperated in this study. The majority was advanced graduate students in psychology. All had successfully completed their first year of graduate study and had passed the departmental Master's Degree Qualifying Examination or its equivalent. In each of the two teaching fellow groups one of the instructor-subjects was a woman. No teaching fellow was an instructor-subject both years. To avoid contamination of the 1960-61 group by knowledge of previous results, no information about 1959-60 results was circulated. There were 17 teaching fellows in the fall of 1960-61 and 18 in the spring of 1960-61.

Student Subjects

All students enrolled in the introductory psychology courses were asked to fill out instructor rating forms. Since it was contrary to department policy to require students to do the ratings if they did not wish to, complete returns could not be secured. Furthermore, not all of the completed forms could be used in our analyses because some students responded in an ambiguous fashion or failed to answer certain items. The ratings of the teaching fellows represent the mean of between 15 and 40 student evaluations collected from all of the sections taught by a given teaching fellow. Usable evaluations forms varied from 18% to 64% of the total enrollments in the semesters reported.

Measures

The peer group nomination form developed by Norman (1961) was administered to the teaching fellows in a teaching seminar they attended. Each teacher was asked to place a third of his fellow teachers at each pole of 20 items defined by "polar descriptions" and covering the five traits referred to earlier. The remaining third on each item were then assigned to a middle group. The score for a teaching fellow on a trait was obtained by subtracting the number of negative pole nominations he received from the number of positive pole nominations he received on the four items defining a given trait.

A self-descriptive adjective checklist, also developed by Norman (1961), was also filled out by the teaching fellows during a meeting of the seminar. This inventory consists of 300 pairs of adjectives equated for social desirability. The respondent is required to indicate which of a pair of adjectives more closely describes him. The five scores obtained from this form correspond to the five traits covered by the peer group nomination form.

In addition each teaching fellow completed Forms A and B of the Institute for Personality and Ability Testing Sixteen Personality Factor (16 PF) Questionnaire (Cattell, Saunders, & Stice, 1957).

The students in the four semesters studied responded to two different evaluation forms. The form used in the year 1959-60 and the fall semester 1960-61 was a short form of 17 questions. The form used in the spring term 1961 consisted of 43 items which were selected to obtain a greater number of teacher-evaluation factors than were obtained on the shorter form. Factor analysis of the short form produced a large general factor. The item loading highest on this factor was:

How would you rate your instructor in general (all-around) teaching ability?

1. An outstanding and stimulating teacher
2. A very good instructor
3. A good instructor
4. An adequate but not stimulating instructor
5. A poor and inadequate instructor

The longer form used in the spring of 1961 contained several items from the shorter form. The item described above was one of these and it loaded highest on the most prominent factor obtained from factor analyzing this new form. Subsequent study of student evaluations with this longer instrument has revealed five consistent factors. In the present paper personality scores of the teaching fellows will be related to the students' evaluation of them on the "overall effectiveness" items, given above, for all semesters and related to the teaching fellow factor scores for the five factors.

RESULTS

The students' responses to the overall teaching effectiveness item were scored 1-5, from most to least desirable, and the means for the various teaching fellows correlated with their scores on each of the 26 personality measures. The correlations are given in Table 1.

The only consistently high relationship is that of Peer Group Nomination Factor 5, Culture, with the students' evaluation. The relation appears in both groups and in all four semesters studied, and leads us to believe that this variable is important. This finding has been replicated at the Ohio State University,[2] where the correlation between peer ratings of culture and student ratings of effectiveness was .48.

Factor scores were obtained for each teaching fellow on the five factors resulting from the analysis of the student rating instrument used in the spring 1961. These factors and the items which loaded highly on them are presented in Table 2. The labels for the factors are somewhat arbitrary, but are intended to reflect what the general character of the item cluster is.

The significant correlations of the personality scores of the teaching fellows with the factor scores for each of the five factors are presented in Table 3.

DISCUSSION

Since the Peer Group Nomination Scale provides the most stable predictor of the students' rating of a teacher's overall effectiveness, it behooves us to look in detail at its "culture" factor and the procedures used to obtain it. The four subscales over which scores are summed to obtain a total Culture score are characterized by the following polar descriptions:

Artistically sensitive	Artistically insensitive
Intellectual	Unreflective, narrow
Polished, refined	Crude, boorish
Imaginative	Simple, direct

Teaching fellows rated by their peers as artistically sensitive, intellectual, and so on, tend to be the ones rated as effective teachers by their students. How should this be interpreted?

---

[2] The authors express their appreciation to D. Wickens who administered the Peer Nomination Form and student rating form at Ohio State University.

First of all we should note that because this is a peer nomination procedure we have no information as to whether the teachers with superior student evaluations are really more artistically sensitive, intellectual, etc. We may conclude only that the other teaching fellows see them as more "cultured" than the less favorably rated members of the group. It remains for further research to determine whether this impression of greater cultural attainment is really based on greater cultural skills.

The students, themselves, tend to see

TABLE 1

CORRELATIONS* BETWEEN PERSONALITY VARIABLES AND STUDENT RATINGS OF OVERALL TEACHING ABILITY

Personality variable	Semester			
	1959-60		1960-61	
	Fall	Spring	Fall	Spring
Peer Group Nominations				
1. Surgency	.40	.41	.50*	.60**
2. Agreeableness	.51*	.64**	.52*	−.01
3. Dependability	.41	.26	−.06	−.11
4. Emotional Stability	.31	.47	.50*	.31
5. Culture	.60*	.65**	.67**	.40
DAI				
1. Surgency	.12	.23	.40	.51*
2. Agreeableness	.07	−.03	.03	−.07
3. Dependability	−.13	−.25	−.35	−.49
4. Emotional Stability	.01	.07	.53*	−.42
5. Culture	.27	.11	.25	.25
Cattell factors				
1. Warmth	.19	.36	.36	.11
2. Brightness (g)	−.06	−.03	−.04	−.27
3. Mature, calm	−.07	−.05	−.60*	−.30
4. Aggressiveness	.42	.50*	.27	.45
5. Enthusiasm	.51*	.56*	.43	.51*
6. Conscientiousness	−.50	−.50*	−.05	−.29
7. Adventurous	.46	.46	.23	.05
8. Sensitivity	.21	.50*	.02	.18
9. Jealous, suspecting	.00	−.03	−.19	−.03
10. Bohemian	.33	.43	−.05	.47
11. Sophisticated	−.31	−.34	−.35	−.42
12. Anxious, insecure	.05	.03	.08	.30
13. Experimenting, critical	.43	.47	.24	.44
14. Self-sufficient	−.17	−.08	−.42	.02
15. Controlled, precise	−.33	−.29	.00	−.44
16. Tense, excitable	−.21	−.16	.31	.24
N	16	16	17	16

* Positive correlations designate high scores on personality variables, as named, going with good student ratings.

* $p < .05$.

** $p < .01$.

TABLE 1

FIVE FACTORS OBTAINED FROM THE STUDENT EVALUATIONS OF THE SPRING OF 1961 AND THE ITEMS WHICH LOADED HIGHEST ON THEM

Factor and Item	Loading	
	Men students	Women students
I. Skill		
Put materials across in interesting way	.862	.847
All-around teaching ability	.818	.809
Stimulated intellectual curiosity	.759	.813
Explained things clearly and concisely	.782	.706
Skillful in observing student reactions	.588	.643
Changed approach to meet new situations	.689	.674
Tried to increase interest of class in subject	.640	.687
Class time well spent	.628	.574
Made clear how each topic fit into course	.561	.471
Overall value of course	.591	.589
Explained why he did things	.556	.642
Anticipated difficulties before they arose	.573	.353
Improved ability to understand deviant individuals	.334	.621
Improved ability to deal with personal problems	.485	.328
Was aware when students failed to keep up	.581	.348
Explained reasons for his criticism	.580	.463
High degree of cultural attainment	.574	.569
II. Overload		
Assigned a great amount of reading	.709	.885
Assigned difficult reading	.670	.850
Asked for more than students could get done	.337	.671
III. Structure		
Planned class activities in detail	.593	.717
Had everything going on schedule	.613	.658
Followed an outline closely	.645	.680
Decided in detail what and how things should be done	.417	.708
IV. Student Rapport		
Invited criticisms of his acts	.463	.270
Told students when they had done a good job	.428	.364
Complimented students in front of others	.424	.490
Able to get personal help in course	.313	.199
Improved ability to deal with personal problems	.184	.322
Criticized poor work	−.306	.223
Was sensitive to students' desire to ask questions	−.310	−.316
Anticipated difficulties before they arose	.283	.171
Permissive and flexible	.073	.328
V. Group Interaction		
Students argued with each other and instructor without hostility	.383	.486
Students frequently volunteered own opinions	.302	.494
Students felt free to express own opinions	.138	.283
Students were friendly	.384	.189
Instructor was aware when students failed to keep up	.423	.012

teachers they judge to be effective as possessing high cultural attainment. An item asking the students to rate the extent to which the instructor exhibited a high degree of cultural attainment correlated .48 with the criterion item of overall teaching effectiveness. Items from the student rating form asking for evaluation of the teachers' talkativeness correlated approximately .36 with the criterion overall effectiveness item, and an item asking for evaluation of the instructors' permissiveness and flexibility correlated only .24 with the same criterion item.

The fact that this appearance of general cultural knowledgeability should relate to the student ratings is not without precedent. For example, Congreve (1930) found students' opinions of the extent to which instructors were "well-read" and informed about material in related fields to load on a factor best labeled as an index of mastery and organized of course content. Beardslee and O'Dowd (1961) also found the student

TABLE 5

Correlations between Personality Variables and Factor Scores Obtained from Student Ratings

Men's student rating factors and personality variable	r	Women's student rating factors and personality variable	r
I. Skill		I. Skill	
PG 1—Surgency	.60	PG 1—Surgency	.53
PG 6—Emotional Stability	.77	PG 4—Emotional Stability	.66
PG 5 Culture	.58	PG 5—Culture	.67
DAI 1—Surgency	.54	DAI 1—Surgency	.66
DAI 1—Dependability	−.50	DAI 4—Emotional Stability	−.60
16 PF (F) Enthusiastic, talkative	.53	16 PF (E)—Dominant, aggressive	.49
16 PF (Q_1)—Experimenting, critical, radical	.50	16 PF (F)—Enthusiastic, talkative	.57
		16 PF (N)—Sophisticated, polished	−.56
II. Overload: No correlation		II. Overload: No correlation	
III. Structure		III. Structure	
16 PF (Q_3)—Lax, unsure	−.68	16 PF (Q_3)—Lax, unsure	−.50
IV. Student Rapport		IV. Student Rapport	
PG 1—Surgency	.53	PG 1—Surgency	.48
DAI 3—Dependability	−.50	PG 4—Emotional Stability	.54
16 PF (Q_1)—Experimenting, critical, radical	.52	PG 5—Culture	.52
V. Group Interaction		V. Group Interaction	
PG 1—Surgency	.51	PG 1—Surgency	.74
PG 5—Culture	.49	PG 4—Emotional Stability	.55
		PG 5—Culture	.63
		DAI 4—Emotional Stability	−.53
		DAI 5—Culture	.49
		16 PF (C)—Mature, calm	−.62
		16 PF (M)—Bohemian, unconcerned	.50
		16 PF (Q_4)—Tense, excitable	.50

Note. Only those correlations which exceed the .01 level of confidence are reported.
Abbreviations: PG, Peer Group Nominations; DAI, Descriptive Adjective Inventory; 16 PF, Cattell factors.

image of college professors emphasized their all-around intellectual competence and sensitivity to artistic and esthetic experience. College professors who overtly manifest these characteristics were perceived by students as closer to the ideal than those who do not.

There may be some question, however, whether this salience of a general culture factor will generalize to success in teaching in areas other than psychology. The work of Riley, Ryan, and Lifshitz (1950) indicates that there may be some differences among what students think are most important for instructors in the arts, the sciences, and the social sciences.

In our study the patterns of correlation for the other variables with the effectiveness item were inconsistent. In three out of the four semesters, however, the Peer Group Nomination variables Agreeableness, and Emotional Stability, and the 16 PF Enthusiasm score correlated well with the student evaluations. We have no hesitancy in saying that a college teacher's possession of agreeableness, emotional stability, and enthusiasm, as well as high cultural attainment, augurs well for him. The preceding statement seems in fact to be more a validation of our measures than a discovery of the attributes of a successful teacher.

Some of the variables that failed to make the grade excite our interest. Do students care whether a teacher is dependable? Would they just as soon have "cold" teachers as "warm" ones? In another study McKeachie (1961) found warmth, as rated by students and observers, to relate to achievement differently for men and women and for students high and low in need affiliation. If warmth affects some students positively and others negatively, the effect may cancel out in a simple correlational study of the sort reported here. Other variables may interact in similar ways.

The Descriptive Adjective Inventory (DAI) scores yielded only two significant correlations with the overall effectiveness item. The tendency for the Emotional Stability scores in 1960-61 to be negatively correlated may be of some moment. Since the correlations for this score evaluated via Peer Group Nominations are positive it may mean that in this group of teachers the ones who were regarded by their peers as most stable were disposed to regard themselves as unstable, and vice versa. If this sort of lack of correspondence is characteristic it may undermine the utility of the self-report instrument. Norman's (1962) results, however, may be interpreted more optimistically, since he found a correlation of .40 between the Emotional Stability scores in the two instruments.

When personality variables are considered in relation to the student rating factors, the same pattern described above for the students' responses to the overall effectiveness item is found for Factor I for both sexes. There are, however, some different personality correlates of the Skill factor for men and for women (see Table 3).

No significant personality correlates were found for Factor II (Overload). The factor analysis of the student rating form did, however, separate this overload factor, which deals with assignments, from one which describes an authoritarian class structure, Factor III. The Cattell personality dimension (Q_3) correlates with this student rating factor in such a way that the person who describes himself as lax and unsure of himself is the one who is compulsive in classroom structure.

Several personality variables correlate with the student rating factor called Student Rapport, but only one, Surgency, appears for students of both sexes. The peer group "surgency" dimension reflects the talkative and enthusiastic individual who seems to be perceived as being able to get along well with the students. It appears as a correlate of the student rating factor Group Interaction also, along with Culture.

We regard the relations we found between the teacher personality variables and the student rating factors as established only tentatively, but in the main they seem reasonable. We have not attempted as yet to argue the desirability of high standing on any of the factors, but we would be inclined to favor Skill, Student Rapport, and Group Interaction. Under these circumstances, teachers high on Surgency, Culture, and Emotional Stability (as judged by their peers, not by themselves) stand out as most likely to be seen in a good light by their students.

REFERENCES

Beardslee, D. C., & O'Dowd, D. D. Higher education and the occupational world. In N. Sanford (Ed.), *The American college.* New York: Wiley, 1962.

Cattell, R. B. Confirmation and clarification of primary personality factors. *Psychometrika,* 1947, 12, 197-220.

Cattell, R. B., Saunders, D. R., & Stice, G. *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

Cosgrove, D. J. Diagnostic rating of teacher performance. *J. educ. Psychol.,* 1959, 50, 200-204.

Della Piana, G. M., & Gage, N. L. Pupils' values and the validity of MTAI. *J. educ. Psychol.,* 1955, 46, 167-178.

Gage, N. L., & Suci, G. Social perception and teacher-pupil relationships. *J. educ. Psychol.,* 1951, 42, 144-152.

Gowan, J. C. Relation of the "K" scale on the MMPI to teaching personality. *Calif. J. educ. Res.,* 1955, 6, 208-212.

McKeachie, W. J. Motivation, teaching methods, and college learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1961.* Lincoln: Univer. Nebraska Press, 1961. Pp. 111-146.

Norman, W. T. Development of self-report tests to measure personality factors identified from peer nominations. *USAF ASD tech. Note,* 1961, No. 61-44.

Norman, W. T. Validation of personality tests as measures of trait-rating factors. *USAF PRL tech. docum. Rep.,* 1962, No. 62-4.

Riley, J. W., Ryan, B. F., & Lifshitz, M. *The student looks at his teacher.* New Brunswick: Rutgers Univer. Press, 1950.

Ryans, D. G. *Characteristics of teachers: Their description, comparison, and appraisal.* Washington, D. C.: American Council on Education, 1960.

Symonds, P. M. Reflections on observations of teachers. *J. educ. Res.,* 1950, 43, 688-696.

Symonds, P. M., & Dudek, S. Use of Rorschach in diagnosis of teacher effectiveness. *J. proj. Tech.,* 1956, 20, 227-234.

Tupes, E. C. Relationships between behavior trait ratings by peers and later officer performance of USAF Officer Candidate School graduates. *USAF PTRC res. Rep.,* 1957, No. 57-125.

Tupes, E. C., & Christal, R. E. Stability of personality trait rating factors obtained under diverse conditions. *USAF WADC tech. Note,* 1958, No. 58-61.

Tupes, E. C., & Christal, R. E. Recurrent personality factors based on trait ratings. *USAF ASD tech. Rep.,* 1961, No. 61-97.

(Received November 9, 1962)

IV-C. STUDENT EVALUATIONS OF RECITATION SECTIONS IN BEGINNING ECONOMICS

Robert L. Isaacson

Any attempt to determine characteristics associated with effective college teaching must be concerned with measures of "effectiveness." One popular measure has been the results of student evaluations of the teachers and the classes. To construct and utilize evaluation instruments soundly, however, the dimensions of teachers' classroom behavior susceptible to student evaluation must be known. Only after this has been done will it be possible to look for correlates of ratings of effective teaching which are stable across teacher populations in different content areas.

In 1964 Isaacson, McKeachie and Milholland reported factor analyses of student evaluations in the introductory psychology course at The University of Michigan in which six factors had consistently appeared over four semesters. One naturally wonders whether or not these same factors, or factors like them, would be obtained from student evaluations using a different set of items in another academic area.

This report deals with a factor analysis of the evaluations made by men and women enrolled in the introductory economics course at The University of Michigan in the spring semester, 1963.

1. METHOD

Student evaluations of the textbooks, lecturer, discussion leader, classroom organization and procedures are customarily obtained from students in the introductory economics course at The University of Michigan. The course lectures are handled by a senior staff member while recitation sections are frequently taught by teaching fellows, who are advanced graduate students in the department. The evaluations are made anonymously by the students and are available to the staff member concerned only after the grades for the term have been submitted.

Through the cooperation of Professor Daniel R. Fusfeld, the student evaluations were made available for the current analysis. The instrument contained 79 items. Some of the items called for short descriptive answers and these were eliminated from our analyses. Items which were concerned with textbooks and other reading assignments and items which referred to the lecturer were also excluded. The 28 items used were related to the written assignments made in the class, tests, the organization and structure of the recitation class, qualities of the teacher, the evaluation of the teacher and the course, and the impact of the course upon the student's beliefs and attitudes. The items are reproduced in Table 1.

While all students filled out evaluation forms, not all of the relevant items were completed in all questionnaires. Other students failed to identify their sex. Only those forms complete in both respects were used. The number of completed forms from the men was 224 and 69 from the women. The imbalance between these numbers reflects a corresponding imbalance in the total course enrollment.

The responses of the students were factor analyzed separately by sex using a principal axes solution followed by Varimax rotation.

2. RESULTS AND DISCUSSION

The loadings of each of the 28 items which were greater than .30 on the factors emerging from the analyses are presented in Table 2. The percent of the extracted variance accounted for by each factor for each sex is given at the bottom of each factor column.

The first five factors in Table 2 appeared to be similar across the sexes. Factor 6 represents a factor apparently distinct for each sex. Since only six factors emerged from the analysis of the men's data, the last four factors of this table represent factors found only in the women's data.

TABLE 1

THE TWENTY EIGHT ITEMS FROM THE ECONOMICS EVALUATION INSTRUMENT WHICH WERE USED IN THIS STUDY

(1) The written papers are:
____1. much too long
____2. somewhat too long
____3. about right length
____4. somewhat too brief
____5. much too brief

(2) The written papers are:
____1. much too complicated
____2. somewhat too complicated
____3. about right
____4. somewhat too simplified
____5. much too simplified

(3) How much do the written papers add to your knowledge of economics?
____1. very much
____2. somewhat
____3. not at all

(4) The presentation:
____1. follows the text much too closely
____2. follows the text somewhat too closely
____3. is about right
____4. strays somewhat too far from the text
____5. strays much too far from the text

(5) The presentation is:
____1. much too complicated
____2. somewhat too complicated
____3. about right
____4. somewhat too simplified
____5. much too simplified

(6) The presentation is organized:
____1. very well
____2. fairly well
____3. fairly poorly
____4. very poorly

(7) The explanations of important points are:
____1. much too detailed
____2. somewhat too detailed
____3. about right
____4. not quite detailed enough
____5. not by far detailed enough

(8) How well does the recitation instructor show how the theory applies to the real world?
____1. very poorly
____2. fairly poorly
____3. fairly well
____4. very well

(9) How much time is left for class discussion and questions?
____1. much too much
____2. somewhat too much
____3. about right amount
____4. somewhat too little
____5. much too little

(10) The tests are:
____1. too many
____2. right in number
____3. too few

(11) The tests are:
____1. too easy
____2. about right
____3. too hard

(12) The tests seem to cover the important things in the course:
____1. very well
____2. fairly well
____3. fairly poorly
____4. very poorly

TABLE 1 (Continued)

13) The marking in this course is generally:
___1. much too easy
___2. somewhat too easy
___3. about right
___4. somewhat too hard
___5. much too hard

14) The instructor's standard of marking appears to be:
___1. very consistent
___2. fairly consistent
___3. fairly inconsistent
___4. very inconsistent

15) His knowledge in the subject matter appears to be:
___1. very deep
___2. fairly deep
___3. fairly shallow
___4. very shallow

16) When he comes to class he appears to be:
___1. very poorly prepared
___2. fairly poorly prepared
___3. fairly well prepared
___4. very well prepared

17) His respect for students as persons appears to be:
___1. very low
___2. fairly low
___3. fairly high
___4. very high

18) His willingness to help students appears to be:
___1. very high
___2. fairly high
___3. fairly low
___4. very low

(19) In matters of opinion (rather than matter of fact or logic) his tolerance of views other than his own appears to be:
____1. very high
____2. fairly high
____3. fairly low
____4. very low

(20) If he is wrong in class, how willing is he to admit it?
____1. very unwilling
____2. fairly unwilling
____3. very willing

(21) Does he try to make you question your own beliefs and ideas in economic matters?
____1. yes, very much
____2. yes, somewhat
____3. yes, little
____4. no, not at all

(22) How do you feel about this? (applies to question 21)
____1. I like it very much
____2. I like it somewhat
____3. It makes no difference to me
____4. I dislike it somewhat
____5. I dislike it very much

(23) Does he overestimate or underestimate the students' level of intelligence?
____1. he overestimates it greatly
____2. he overestimates it somewhat
____3. he estimates it correctly
____4. he underestimates it somewhat
____5. he underestimates it greatly

TABLE 1 (Concluded)

(24) Would you want to take another course from this instructor?

____1. I certainly would
____2. I would not mind
____3. I would prefer not to
____4. I would refuse to

(25) How much do you feel your ideas and beliefs in economic matter have changed since you began to study economics?

____1. very little
____2. fairly little
____3. fairly much
____4. very much

(26) Do you think this course is valuable to you?

____1. it is very valuable for me
____2. it is fairly valuable for me
____3. it has little value for me
____4. it has no value for me

(27) Taking everything into account, how do you feel about this course?

____1. I dislike it very much
____2. I dislike it somewhat
____3. I neither like it nor dislike it
____4. I like it somewhat
____5. I like it very much

(28) How does the recitation section instructor in this course compare with the instructors in courses outside economics which you are taking this term?

____1. very favorably
____2. fairly favorably
____3. neither favorably, nor unfavorably
____4. fairly unfavorably
____5. very unfavorably

TABLE 2

THE LOADINGS OF .30 OR MORE OF STUDENT RATING FORM ITEMS ON THE SIX MALE FACTORS AND TEN FEMALE FACTORS

(The men's loadings are reported in the columns marked "M" and the women's loadings are reported in the columns marked "W." Decimal points omitted. The percent of extracted variance account for by each factor is given at the bottom of each column.)

	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5		Factor 6		Factor 7	Factor 8	Factor 9	Factor 10
	M	W	M	W	M	W	M	W	M	W	M	W	W	W	W	W
1									-50	-55						
2									-56	-50		-38				
3					-37							37				
4							39				-33				-61	
5										-38	37					
6	80	81														
7										-62	-30					
8	-55	-57						36								
9						30	-39	-49							36	
10																31
11											-37			55		
12					-30	-37										32
13												60				
14	36		-42													55
15	48					-39					33			-58		
16	-73	-87														
17	-33		65	78												
18	43		-61	-46								50				
19			-59	-58												
20			52	72												
21							-69	-82								
22	71						-66	-78								
23				32											-30	
24		76	-31													
25					56	81										
26					-71	-39							-46			
27	-32				48								73			
28	81	76														
Percent variance removed	34.6	21.2	19.2	14.0	14.8	9.3	14.5	1.50	9.1	9.0	7.8	7.3	6.9	6.5	5.6	5.2

From the nature of the items loading on the factors we believed that the following descriptive names could be applied to the first five factors.

Factor 1: Skill

The items loading on this factor dealt with the students' opinion of the instructor's organization of the class (Item 6) and the application of economic theory to real world problems (Item 8). Another item (16), was related to how well prepared the teacher was for class. Item 28 asked for the students' opinions, from favorable to unfavorable, of the instructor. The direction of correlations was such that the greater the preparedness of the teacher, the more favorable were the students' reactions toward the instructor. This factor would appear to be most closely related to the "skill factor," (Factor 1), in our earlier report (Isaacson, et al., 1964). It represented a cluster of items related to the overall effectiveness of the teacher.

Factor 2: Student Rapport

Items which loaded on this factor were ones which indicated a teachers' high respect for the students (Item 17) and high willingness to help the students (Item 18) high tolerance for the expression of student opinion (Item 19) and willingness to admit being wrong (Item 20). Along with these items was one which indicated that the teacher maintained a consistent set of standards of evaluating the students. This factor which seemed to be similar to a factor in the analysis of the evaluations obtained from psychology students. In our previous work (Isaacson, et al., 1964) we referred to this factor as "student rapport" and it was one in which that had items relating to the ability of psychology teachers to accept and invite criticism of their acts and to provide personal help in the course.

Factor 3: Value of the Course

The third factor obtained from the evaluations was one which had three items loading highly on it. These were related to the students' change in attitudes and ideas about the course (Item 25) and the value of the course (Item 26), and whether or not the tests covered important points in the course (Item 12). Students who reported a considerable change in their ideas and beliefs about economics thought the course valuable. There is no way to compare this factor with results of our earlier work since comparable items were not present in both studies. The items related to course value loaded on the teacher skill factor in our earlier report of the analyses of student evaluations in the psychology course.

Factor 4: Change in Beliefs

This factor represents an interesting cluster of items which again is not directly comparable to our earlier factors due to lack of corresponding items. Overall, the factor has to do with the change in beliefs of students taking the economics course. The highest loading of the cluster were Items 21 and 22. In these the student was asked whether his experiences in the course made him question his own beliefs and whether or not he liked to have his beliefs questioned. There was a correlation of .68 between these two items. This would suggest that students who had their attitudes changed liked the experience or that those who liked to have their attitudes changed liked the teachers who acted in such a manner as to make them change or question their own structure beliefs. The other item loading on this factor was one in which students reported too little time was allowed for question and answer sessions between the teacher and the students.

Factor 5: Overload

The descriptive title used for this factor is drawn from our previous work and indicates our belief that this factor represents the same dimension observed in the psychology course. In this case the factor represents the degree to which the student finds the work excessive. In the present analysis this classroom dimension is defined by two items describing the length and complexity of the written assignments (Items 1 and 2).

Factor 6 (men): Structure

For the men, at least, another of our earlier factors (Structure) seems to be present. This factor deals with the degree of organization exhibited by the teacher in presenting his material.

This last factor obtained from the analyses of the men's rating was one of four items which indicated that the instructor had a too complicated presentation in class (Item 5) and strayed from the topic in class (Item 4). In addition, Item 11, which asked about the difficulty of the test, and Item 15 went together on this factor. Students who felt that the recitation presentation was too complicated, also felt that the tests were too hard but that the teacher had a deep knowledge of the material he was presenting. Part of this "structure complex" appears in the eighth factor removed from the women's data (Items 11 and 15).

Factor 6 (women): Considerateness

This factor, unique to women, represents a set of items which appear related to the considerateness, or thoughtfulness of the teacher. The two items defining it are 13, dealing with the marking in the course, and 18, asking about the willingness of the instructor to help the student. Item 2, asking whether the written papers are too complicated also appears on this factor, and apparently brings along with it Item 3, on whether the written papers contribute to knowledge of economics.

The seventh factor obtained from the women's data shows a doublet of items consisting of Items 26 and 27 which have to do with the value of the course and whether or not the students liked it. The significance of the separation of Factor 7 from Factor 3 is not clear, except that it may reflect some greater discriminative ability of the women in these areas covered by three items, 25, 26, and 27.

Factor 8 from the factor analysis of the women's data is a doublet of two items (Items 11 and 15), in which the teacher is marked as being too easy and having shallow knowledge in the field of economics. These items were found to appear in male Factor 6.

The ninth factor consists of three items (Items 4, 9, and 23) which collectively suggest a teacher who underestimates student abilities, strays from the topic, and perhaps as a result, leaves too much time for discussion.

The tenth factor represents a cluster of three items (Items 10, 12 and 14) which have to do with the teacher's testing and grading policies. One extreme of the dimension represents an ability to devise tests which test important facets of the course, to be consistent in grading, and a tendency to have too many tests.

3. CONCLUSION

Even though the student evaluation form used in the economics course was made up of different items than were used in our previous study of psychology students (Isaacson, et al., 1964), four factors from this earlier work seem to be represented. These are: Teacher Skill, Student Rapport, Overload, and Structure. The economics students of both sexes also exhibited two factors which were not previously found in the psychology course. One was related to their change in attitudes and beliefs about economics (Change in Beliefs) and another concerned with the value of the Course (Value of the Course). The fact that four of our six "stable factors" were present in the present study despite the use of a different form and in another academic area encourages the belief that they may approach fundamental dimensions of classroom instruction. The failure to find the other two factors "Evaluation" and "Interaction" may be due to a lack of items adequately testing these dimensions in the current questionnaire.

The factors related to the changing of attitudes and beliefs found in the data from the economics students suggests that items like those loading on this factor should be included in future evaluation instruments so that its stability across sexes, instructors, and academic areas may be ascertained.

Reference

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, Margaret, Baerwaldt, J. W., and Zinn, K. The dimensions of student evaluations of teaching. In preparation, 1964.

Student Characteristics Related to Achievement (Main Effects,

Yi-Guang Lin

Four categories of student characteristics related to performance were studied. The first category is intellectual ability. In the 1961 sample, this was measured by scores on the American Council on Education Psychological Examination (ACE). In the 1963 sample, the College Entrance Examination Board Scholastic Aptitude Test (SAT) was used.

The second category includes study habits and skills. Twenty items were selected from the Brown-Holtzman Survey of Study Habits and Attitudes (SSHA)[1] (Brown-Holtzman, 1955). The items selected were those most relevant to the student's habits and skills in utilizing class time, in preparing class notes and examinations, and in integrating the subject matter.

The third category is student personality attributes. Cattell's Sixteen Personality Factor Questionnaire (16PF) Form A (Cattell, 1956) was used in the 1961 study. Five scores of Fricke's Opinion, Attitude and Interest Survey (OAIS) Fricke, 1963): Achiever Personality (AP), Intellectual Quality (IQ), Creative Personality (CP), Social Adjustment (SA), and Social Science Interest (SSI) were used in both the 1961 and 1963 studies.

The fourth category is achievement anxiety. The measure used was the Alpert-Haber Achievement Anxiety Test (Alpert and Haber, 1960). The test consists of two subscales, one for debilitating anxiety (DA) and one for facilitating anxiety (FA). Several studies have indicated that it is useful to distinguish the detrimental and facilitating aspects of test anxiety and their effects on performance (Alpert and Haber, 1960; Reese, 1961; Ruebush, 1960; Sarason, I., 1957, 1959, 1961; Sarason and Mandler, 1952).

Subjects

The subjects were students enrolled in introductory psychology courses at the University of Michigan in the Spring semesters of 1961 and 1963.

---

[1] Permission has been obtained from Psychological Corporation to use this measure for research purposes.

Performance Measures and Method of Analysis of Results

As measures of achievement, we used course grade, the Criteria test, and a short test of content knowledge. Both the Criteria test and the content test were given at the final examination as parts of the total evaluation of students' learning and achievement in the course.

In 1961 the content test consisted of 15 multiple-choice items; in 1963 a different test was used. It had 25 items, and was given only in the Psychology 101 course. Because of these vagaries it was decided to look for the presence of significant mean differences between groups that were above and below the median on each student characteristic before proceeding to a more elaborate analysis. In only ten of the 36 comparisons were the differences statistically significant, four of them occurring in the case of measures of intellectual ability, two more on the Intellectual Quality Scale of the OAIS. The results are shown in Table 1, and seemed to add little to what appeared for grades and Criteria Test. No further analysis of these results was carried out, therefore.

For the Criteria Test and course grades, the measure used to show relationship between student characteristics and performance measures was the correlation coefficient. Separate analyses were made for the two sexes.

Results and Discussion

The correlation for measures which were available both years are shown in Table 2. The results reported in this section deal only with the main effects on performance.

Intellectual Ability

Intellectual ability as measured by ACE and SAT has shown a very significant main effect on performance in both the 1961 and 1963 studies. The effect seems to be more pronounced for the Criteria test than for the course grade.

Study Habits and Skills

Study habits and skills showed no significant main effect in the 1963 study, but it did appear in three of the four groups in the 1961 study. Correlations were generally low, however.

Table 1

Significance Levels for Comparisons of Mean Scores on Content Knowledge Tests for Students in Introductory Psychology Who Were Above the Median or Below the Median on Various Measures

Measures on Which Class was Dichotomized at the Median	Significance Level of Mean Difference on Content Test			
	1961 (Psychology 31)		1963 (Psychology 101)	
	Men	Women	Men	Women
ACE	.001	.001		
SAT			.01	.001
SSHA	NS*	NS	NS	NS
AP	.02	NS	NS	NS
IQ	NS	.01	NS	.01
SA	NS	NS	NS	NS
CP	.05	NS	NS	.05
SSI	NS	NS	NS	NS
DA	NS	NS	NS	.01
FA	NS	NS	NS	.02

* Not significant

Table 2

Correlation Between Various Student Attributes and Course Grades or Criteria Test Scores

	Correlation * with Course Grade				Correlation * with Criteria Test Score			
Measure	1961		1963		1961		1963	
	Men	Women	Men	Women	Men	Women	Men	Women
ACE	41	33			41	42		
SAT			36	41			58	55
SSHA	23	15	10	16	10	-03	01	00
AP	21	14	18	17	14	18	21	11
IQ	14	21	21	27	30	42	32	42
SA	11	-06	02	04	03	-04	-01	-05
CP	09	11	00	09	17	11	00	20
SSI	15	17	03	22	16	10	03	22
DA	-12	-18	-20	-10	-12	-22	-22	-19
FA	15	18	18	14	06	24	28	15
N	202	284	218	279	202	284	218	279

* r required for significance at .05 level (2-tail),

for N = 200: .14
for N = 300: .11

16 PF Variables

Table 3 presents the correlation coefficients between Cattell's 16 PF variables and the three performance measures for the male and the female groups in the 1961 study. The coefficients are in general very low, the only correlations as large as .2 being those between intelligence and the outcome measures.

OAIS

Out of the five scores of Fricke's OAIS used in this study, the Intellectual Quality (IQ) scale has shown the most significant and consistent main effect on performance measures in both 1961 and 1963 studies (Table 2). The Intellectual Quality scale is designed to measure those personality attributes associated with intelligent behavior and an intellectual orientation. It might be considered a scale to measure intelligence through the use of non-intellectual material. The fact that intellectual orientation as a personality attribute is outstanding tends to enhance further the importance of intellectual abilities in determining learning.

Achiever Personality also correlated significantly with grades and scores, but not so closely as Intellectual Quality. The other OAIS scores showed only scattered significant correlations.

Test Anxiety

Correlations for Facilitating Anxiety and Debilitating Anxiety were in the expected directions (positive for FA, negative for DA) but rather low.

Regression Analysis Results

Because correlations exist among some of the independent variables such as ACE, debilitating anxiety, facilitating anxiety, and intellectual quality the main effects of these variables are not independent. Eight regression analyses were therefore performed using course grades and Criteria test scores as dependent variables to find how the variables would combine to affect the performance measures.

Table 4 shows the regression analysis results. As expected, the intellectual ability variable appears in all regression equations and it accounts for most of the predictable variance. Intellectual Quality appears in all equations for predicting Criteria Test Scores, but in none for predicting grades. On the other hand, Social Science Interest and Study Skills each appears among the grade predictors but not in the Test predictors. Except for the 1961 sample of men, Criteria Test scores are more predictable than grades.

Table 3

Correlations* Between Cattell's 16 PF and Performance Measures for the Men and Women in the 1961 Study

Variables	Course Grade		Criteria Test	
	Men	Women	Men	Women
A: Warmth	-.098	-.030	-.026	-.103
B: Intelligence	.079	.158	.222	.366
C: Emotional stability	-.019	.013	-.052	.056
E: Dominance	.007	.104	.112	.027
F: Surgency	-.175	-.114	-.122	-.130
G: Conscientiousness	-.059	-.027	-.073	-.119
H: Adventurous	-.035	-.046	.034	.164
I: Sensitivity	.000	.041	.109	.139
L: Paranoid tendency	-.021	.018	.006	-.160
M: Bohemian	.071	.098	.055	.011
N: Shrewdness	-.061	.093	-.074	-.026
O: Anxious, Insecure	-.167	-.027	-.098	.011
Q_1 : Radicalism	.148	.185	.127	.183
Q_2: Self-sufficient	.173	.088	.199	179
Q_3: Controlled	.028	.134	-.011	.051
Q_4: Tense, excitable	-.066	-.141	-.075	-.195
N	251	313	237	312

*r required for significance at the .05 level (2-tail)
for: N = 200: .14
for: N = 300: .11

Table 4

Significant Variables Entering into the Regression Equations

I. Course Grade as Outcome Measure

Group	Predictor Variables and Zero order r's	Multiple R
1961, Male (N=202)	ACE (.411), SSHA (.231), AP (.211)	.488
1961, Female (N=284)	ACE (.326), SSHA (.152), SSI (.171)	.392
1963, Male (N=218)	SAT (.355), AP (.185)	.392
1963, Female (N=279)	SAT (.411), SSI (.223), SSHA (.163)	.465

II. Criteria Test as Outcome Measure

Group	Predictor Variables and Zero order r's	Multiple R
1961, Male (N=202)	ACE (.406), IQ (.304)	.447
1961, Female (N=284)	ACE (.422), IQ (.421), AP (.182) FA(.241)	.552
1963, Male (N=218)	SAT (.578), AP (.210), IQ (.325)	.624
1963, Female (N=279)	SAT (.548), IQ (.419)	.578

Summary

The present study describes the effect of each variable or performance measure on achievement as it stands, without an attempt to control other variables at the same time. To the extent that the compared groups are balanced and the relevant independent variables are not correlated, the comparison of mean performances between groups might indicate the pure effect of each independent variable. Otherwise, the main effect might be covered up or reduced.

The results of the two studies of main effects may be summarized as follows:

1. Of the variables we have investigated in this study, intellectual ability is the most important and the best predictor variable for the outcome performances as measured by the course grade and the Criteria test. Intellectual ability is more closely related to Criteria test performance than to course grade.

2. Some non-intellectual student characteristics seem to be related to performance in the introductory psychology course. These are study habits and skills, academic motivation, and interest in social science, but they seem to be more important in determining course grade than performance on the Criteria test.

3. Several scales of Cattell's 16 PF questionnaire were found to correlate significantly with course grade and Criteria test measure. The coefficients, however, are generally very low.

4. Considerable variance in performance is left unaccounted for when intellectual ability is the predictor. Including non-intellectual variables increases predictability only slightly.

References

Alpert, R., and Haber, R. W. Anxiety in academic achievement situations. J. abnorm. soc. Psychol., 1960, 61, 207-215.

Brown, W. F., and Holtzman, W. H. A Study-Attitude Questionnaire for predicting academic success. J. educ. Psychol., 1955, 46, 75-84.

Cattell, R. B. Sixteen Personality Factor Questionnaire. Form A. Champaign, Illinois: Institute for Personality and Ability Testing, 1956.

Fricke, B. G. Opinion, Attitude and Interest Survey Handbook. Ann Arbor: Evaluation and Examinations Division, University of Michigan, 1963.

Reese, H. W. Manifest anxiety and achievement test performance. J. educ. Psychol., 1961, 52, 132-135.

Ruebush, B. K. Interfering and facilitating effects of test anxiety. J. abnorm. soc. Psychol., 1960, 60, 205-212.

Sarason, I . G. Test anxiety, general anxiety, and intellectual performance. J. consult. Psychol., 1957, 21, 485-490.

Sarason, I . G. Intellectual and personality correlates of test anxiety. J. abnorm. soc. Psychol., 1959, 59, 272-275.

Sarason, I . G. Test anxiety and the intellectual performance of college students. J. educ. Psychol., 1961, 52, 201-206.

Sarason, S. B., and Mandler, G. Some correlates of test anxiety. J. abnorm. soc. Psychol., 1952, 47, 810-817.

V-B. RELATIONS BETWEEN n. ACHIEVEMENT, n. AFFILIATION, DEBILITATING ANXIETY, AND SEVERAL MEASURES OF COLLEGIATE PERFORMANCE

Robert L. Isaacson, Wilbert J. McKeachie,
John E. Milholland, and Yi-Guang Lin

At face value the college envirionment would seem to provide an archetype of a competitive situation. Students compete for grades with considerable intensity. Cues suggesting opportunities for competition with standards of excellence arise from classroom organization, the instructors, the other students, and competition for grades is part of students' expectations.

Yet, it is by no means clear that the motive for achievement (McClelland, Atkinson, Clark, and Lowell, 1953) is unequivocally related to competition with academic standards of excellence. Lowell (1952) reported a positive correlation between n. Achievement scores and academic performance, but in general such correlations have been "at best, low to moderate" (Atkinson, 1958, p. 605). This conclusion was based on the findings of Morgan (1953) and Ricciuti and Sadacca (1955). Attempts to find a relationship between grades in the introductory psychology course at The University of Michigan and n. Achievement scores, controlling the possible interactive effects of intelligence by examining differential effects of the motive with several regions of the distribution of ACE scores, were unsuccessful (Isaacson and Patricia O'Connor, unpublished data) and Fricke has failed to find any first order relationship between n. Achievement scores and total grade point averages attained at The University of Michigan (Fricke, 1962).

The lack of relationship may be due to the fact that academic performance is highly overdetermined. How well a student performs is affected by many motives. With the development of tests for measuring motives other than n. Achievement it becomes theoretically possible to take into account interactions among motives. However, still other factors may act to diminish the strength of relationships found between the achievement motive and collegiate performance.

Atkinson (1958) has distinguished between motive and motivation. According to this theory, the total motivation toward a given task is the product of (1) the latent motive for achievement (Ms) as measured in the manner described by McClelland, et al. (1953), (2) the subjective probability of goal attainment (Ps), and (3) the incentive value of the goal (Is). This last variable (Is) is defined as (1-Ps). The model predicts, for example, that people high in the motive of n. Achievement would work most diligently when the value of Ps was of intermediate value, because it is at this point that motivation (the product of Ms, Ps, Is) is greatest. When the Ps is too high or too low, motivation will be less than that at the intermediate Ps values. This interaction would diminish the overall predictive value of n. Achievement.

In a similar fashion the motive to avoid failure can be distinguished from the total motivation to avoid failure. The total motivation to avoid failure can be expressed as the product of probability of failure (Pf) the latent motive to avoid failure (Maf), and the negative incentive value of failure (If). When proposing this motivational theory, Atkinson and Litwin (1960) suggested that failure avoidance can be measured by an objective measure of anxiety, such as the Taylor Manifest Anxiety Scale (1953) or the Mandler-Sarason Test Anxiety Questionnaire (1960). In the research to be reported here, the Alpert-Haber Anxiety Scale (1960) has been used as an index of the latent motive to avoid failure.

Measures of "anxiety" have been frequently found to have negative relationships with college performance. Spielberger (1962) reported a curvilinear relationship between anxiety scores, intellectual brightness levels, and performance. The students whose performance was most seriously impaired by anxiety were those in the middle range of intellectual ability, a finding which might be predicted on the basis of Atkinson's formulation, if the middle ability range is that in which uncertainty about success or failure is greatest.

A measure of a motive for the establishment of warm, friendly relations with others (n. Affiliation) has been developed (Shipley and Veroff, 1952). Scores representing strength of this affiliative motive have not directly related to academic performance, however, probably because a relationship between the affiliation motive and college academic performance is not demanded by theoretical considerations. Nevertheless, it may be that whenever affiliative goals can be reached by doing well in scholastic endeavors, individuals high in the affiliation motive will perform well. In most colleges certain levels of academic performance are prerequisite to certain affiliative satisfactions e.g., joining a fraternity, participating in extracurricular activities.

As part of our general research efforts directed toward understanding effective college teaching, measures of n. Achievement, n. Affiliation, and scores from Achievement Anxiety tests were obtained for students enrolled in the introductory psychology course at Michigan in the Spring of 1961. Scores on the American Council on Education Psychological Examination (ACE) were also available. Several analyses were conducted in the attempt to assess the relationship between these variables and five measures of academic performance.

1. TEST ADMINISTRATION

Scores for freshman on the ACE were obtained from the Evaluation and Examinations Division of the University, since all entering freshmen had been required to take the test. Scores were not always available for students transferring into the University with advanced credit from other institutions, however.

The scores for n. Achievement and n. Affiliation were obtained from protocols written in response to pictures, according to the method described by McClelland, Atkinson, Clark, and Lowell (1953) and Smith and Feld (1958). Students were tested in their usual classrooms, although their teachers were not present. Research assistants administered the projective tests separately to men and women. Usually the men assembled at one end of the classroom while the women were at the other. The instructions and atmosphere were "neutral." Students were provided forms to write upon and each had a booklet of pictures. The pictures were the same as those used by Veroff, Atkinson, Feld, and Gurin (1960). The testing sessions for the projective devices extended over the first half of the Spring Semester, 1961. The protocols were scored by trained scorers with reliabilities over .90 established with expert scorers.

The test of debilitating and facilitating anxiety related to achievement was administered in a special testing session in the evening. Men and women were tested separately in groups of from 100 to 150 students.

2. TRANSFORMED ACHIEVEMENT-ANXIETY SCORES

In some of the subsequent analyses raw scores were used, but for others combined n. Achievement and debilitating anxiety scores were obtained by subtracting each student's standard score on anxiety from his standard score on n. Achievement. Students with positive "difference scores" thus have a relatively higher position in the n. Achievement distribution than in the anxiety distribution, and students with negative "difference scores" are higher on anxiety than on n. Achievement. (For more details of this technique see Isaacson, 1964, in press.) Our interpretation of these scores is that the positive scores reflect relatively greater achievement or "success" orientation whereas negative scores reflect relatively greater fear of failure.

3. ACADEMIC PERFORMANCE MEASURES

Five measures of academic performance were studied. These were: (1) grade in the course, (2) cumulative grade point average at The University of Michigan through the spring term, 1961, (3) the score obtained by the students on a "Criteria Test,"* (4) a Residual Criteria Test score consisting of the difference between an individual's obtained score and his score predicted from his ACE score,** and (5) the score obtained by the students on a special Fifteen-Item Test of psychological content administered as part of the final examination

---

*Introductory Psychology Criteria Test, developed jointly by the project staff and by the Committee on Criteria for the First Course of Division 2 (Teaching) of the American Psychological Association.

**The Residual Criteria Test Score was the error of estimate obtained when a student's Criteria Test Score was predicted from his ACE score. A transformation of scores was made to produce a distribution of residual scores with a mean of 500 and a standard deviation of 10.

4. RESULTS

Overall Correlations With Academic Performance

The raw motive scores obtained for the students were correlated with the five performance scores. The results obtained from the men are presented in Table 1 and those obtained from the women in Table 2.

The only correlations of personality measures with academic performance which reach an acceptable level of significance are those of the two measures of anxiety. Correlations of academic performance with n. Achievement and n. Affiliation do not suggest any relationship.

Several other analyses of the relations of personality measures to student performance were made, using the men in our sample. The sex restriction was imposed because most of the positive results with n. Achievement have been limited to men. The nature of achievement motivation in women is not well understood.

As a preliminary step, the men were first divided into "success-oriented" and "failure-oriented" groups in accordance with their relative positions on n. Achievement and on Anxiety, as previously described. A comparison of the ACE scores for the two groups is shown in Table 3. It indicates that the difference in average ACE scores between the "success-oriented" and "failure-oriented" groups within each category of ability was negligible.

Theory would predict that a greater preformance superiority of high Achievement Motivation groups would appear in the middle range of intelligence level than at other ranges of ability if it is assumed that students of intermediate ability would perceive themselves to have intermediate probabilities of success (Ps).

Table 4 presents the relationships to our five performance measures of the achievement motive and test anxiety for men in the various ACE categories. Statistical significance test (t-tests) were applied only to the differences between students in the middle range ability groups.

Contrary to prediction, the failure-oriented motive group in the middle ability range showed a higher performance level on all five measures. Three of the five differences were statistically significant (overall grade point average, $p < .025$, course grade in introductory psychology $p < .025$, 15-Item Test of knowledge in psychology $p < .005$). Since all the data refer to one

TABLE 1

PRODUCT MOMENT CORRELATIONS BETWEEN MOTIVE SCORES AND THE FIVE PERFORMANCE MEASURES FOR MEN STUDENTS

Motivation Measures	Course Grade in Psychology	Grade Point Average	Criteria Test Score	Residual Criteria Test Score	15-Item Test Score
n. Achievement	.04	-.04	.00	-.02	.03
n. Affiliation	.06	.02	-.05	-.09	-.05
Debilitating Anxiety	-.18*	-.24*	-.17	-.02	-.01
Facilitating Anxiety	.25*	.29*	.12	.01	.05
N	241	252	267	267	193

*Significant at the .05 level (2-tail).

TABLE 2

PRODUCT MOMENT CORRELATIONS BETWEEN MOTIVE SCORES AND THE FIVE PERFORMANCE MEASURES FOR WOMEN STUDENTS

Motivation Measures	Course Grade in Psychology	Grade Point Average	Criteria Test Score	Residual Criteria Test Score	15-Item Test Score
n. Achievement	-.07	-.01	-.10	-.09	-.03
n. Affilation	.04	.03	.04	.01	-.01
Debilitating Anxiety	-.21*	-.22*	-.22*	-.04	-.05
Facilitating Anxiety	.21*	.38*	.19*	.09	.00
N	302	294	366	366	244

*Represents a correlation at which $p < .05$.

TABLE 3

MEAN ACE SCORE OF MEN IN EACH MOTIVATIONAL CATEGORY

FOR EACH OF FIVE ABILITY CATEGORIES

Motivational Category		ACE Score Range					Total
		113 & Below	114-122	123-130	131-141	142 & Higher	
Success Oriented:	Mean	120.8	117.7	126.0	136.2	151.0	129.60
	N	25	20	23	33	36	137
Failure Oriented:	Mean	101.2	118.6	126.3	135.6	150.6	122.87
	N	40	32	31	25	21	149

TABLE 4

PERFORMANCE ON FIVE MEASURES OF ACADEMIC ACHIEVEMENT OF SUCCESS-ORIENTED AND FAILURE-ORIENTED STUDENTS DIVIDED INTO FIVE CATEGORIES OF ABILITY BASED ON ACE SCORES

Measure	ACE Category											Total
	N	≤ 113	N	114-122	N	123-130	N	131-141	N	≥ 142	N	
Cumulative GPA												
Success-oriented	22	2.262	16	2.301	22	2.372	30	2.617	29	2.688	119	2.481
Failure-oriented	28	2.058	27	2.156	24	2.653	19	2.555	21	2.573	119	2.371
Difference		.204		.146		-.281**		.062		.115		
Residual Criteria Test												
Success-oriented	25	502.4	19	504.8	21	512.3	33	504.1	35	504.1	113	505.29
Failure-oriented	38	495.7	32	495.6	31	523.9	24	515.5	20	512.0	145	507.21
Difference		6.7		9.2		-11.6		-11.4		-7.9		
Criteria Test (Z Score Units)												
Success-oriented	25	-.539	19	-.096	21	.173	33	.357	35	.645	133	.170
Failure-oriented	38	-.589	32	-.179	31	.292	24	.449	20	.747	145	.046
Difference		.050		.083		-.119*		-.092		-.102		
Introductory Psychology Course Grade												
Success-oriented	24	2.375	19	2.526	22	2.364	33	2.848	34	3.088	132	2.697
Failure-oriented	32	1.906	30	2.300	30	2.833	23	3.043	21	3.286	136	2.603
Difference		.469		.226		-.469**		-.195		-.198		
Fifteen Item Test												
Success-oriented	18	8.333	15	8.800	20	8.050	29	8.793	27	9.963	109	8.872
Failure-oriented	28	8.107	24	8.042	25	9.800	19	9.421	15	9.467	111	8.883
Difference		.226		.758		-1.750**		-.628		.496		

*Variances significantly different.

**Difference significant at .025 level (2-tail).

sample, however, the different tests are not independent, and the stated probability levels may be somewhat in error.

Table 4 also shows the differential effects of achievement motivation on performance at the different ability levels. At the lower range of ability, the high achievement motivation groups showed a tendency toward higher performance on all measures. Again, this is contrary to prediction by theory.

At the higher ability levels there seems to be no clear tendency for one motive group to surpass the other. In general there is no suggestion of superior performance on the part of the success-oriented students in our data.

In Table 5 are comparisons of the performance levels of the high and low n. Affiliation groups (median split) for the five categories of intellectual ability. For the middle range ability groups, the high n. Affiliation group showed statistically significant superior performance on the Criteria Test ($p < .025$) and on the Residual Criteria Test Score ($p < .05$). The high n. Affiliation group also has higher overall grade point average than the low n. Affiliation group, although this difference only reaches the 10% level of significance.

The differential effect of the affiliation motive on performance of different ability groups was suggested in some of these results and as, a consequence, no main effect of the affiliation motive on performance over the whole range of ability groups and performance measures was observed. In total group however, high n. Affiliation students had better GPA's and course grades (subjective measures) but were worse on objective measures. There is a suggestion that high n. Affiliation students do better than low n. Affiliation students in the lowest ability ranges on all achievement measures, but that they do worse than the low n. Affiliation students in the highest ability ranges (excluding the measures of grade).

Joint Effects of n. Affiliation and n. Achievement on Performance

To determine the joint effects of the motives for affiliation, achievement, and debilitating anxiety, the success-oriented and failure-oriented student groups were further subdivided on the basis of whether their n. Affiliation score was above or below the median score of the n. Affiliation distribution. The performance of the motive groups thus produced are given in Table 6.

At the lowest ACE range (ACE less than 114), the group high in n. Affiliation and in the success-orientation outperform all other motive groups on all five dependent measures. This effect is most pronounced with the course grade and cumulative grade point average. The group of success-oriented students who are low in n. Affiliation are generally indistinguishable from the failure-oriented groups at this ability level.

TABLE 5

THE PERFORMANCE ON FIVE MEASURES OF ACADEMIC ACHIEVEMENT OF STUDENTS ABOVE (HIGH AFF.) AND BELOW (LOW AFF.) THE MEDIAN n. AFFILIATION SCORE

Measure	ACE Category											Total
	N	≤ 113	N	114-122	N	123-130	N	131-141	N	≥ 142	N	
Cumulative GPA												
High Affiliation	27	2.209	23	2.257	24	2.614	29	2.609	25	2.673	128	2.475
Low Affiliation	23	2.076	20	2.156	22	2.414	20	2.620	25	2.606	110	2.378
Difference		.133		.101		.200		-.011		.067		
						p < .10						
Residual Criteria Test												
High Affiliation	33	501.2	26	486.5	27	541.3	32	501.8	28	488.9	146	503.8
Low Affiliation	30	495.2	25	512.0	25	495.3	25	518.2	27	526.2	132	509.1
Difference		.60		-25.5		46.0		-16.4		-37.3		
						p < .05						
Criteria Test: Post-test (Z Score Units)												
High Affiliation	33	-.475	26	-.276	27	.485	32	.330	28	.507	146	.103
Low Affiliation	30	-.672	25	-.015	25	-.016	25	.481	27	.864	132	.109
Difference		.197		-.261		.501		-.151		-.357		
						p < .025						
Introductory Psychology Course Grade												
High Affiliation	29	2.207	26	2.269	27	2.630	32	2.969	28	3.179	142	2.662
Low Affiliation	27	2.000	23	2.522	25	2.640	24	2.875	27	3.148	126	2.635
Difference		.207		-.253		-.010		.094		.031		
						n.s.						
Fifteen Item Test												
High Affiliation	25	8.32	22	8.27	24	9.25	29	9.21	19	9.26	119	8.865
Low Affiliation	21	8.05	17	8.41	21	8.76	19	8.79	23	10.22	101	8.892
Difference		.27		-.14		.49		.42		-.96		
						n.s.						

TABLE 6

MEAN PERFORMANCE OF FIVE ACHIEVEMENT MEASURES BY DIFFERENT MOTIVATIONAL GROUPS

Performance Measure	Motivational Orientation	n. Affiliation Position	ACE Range				
			113 & Lower	114-122	123-130	131-141	142 & Higher
Cumulative GPA	Success	Above Median	2.440	2.415	2.447	2.639	2.725
		Below Median	2.093	2.135	2.730	2.451	2.607
	Failure	Above Median	2.138	2.110	2.266	2.579	2.653
		Below Median	1.994	2.175	2.561	2.670	2.536
Grade in Course	Success	Above Median	2.636	2.545	2.200	2.905	3.176
		Below Median	1.944	2.067	2.882	2.901	3.182
	Failure	Above Median	2.154	2.500	2.500	2.750	3.000
		Below Median	1.857	2.533	2.769	3.000	3.400
Criteria Test Score	Success	Above Median	-.425	-.014	.117	.375	.508
		Below Median	-.500	-.468	.668	.243	.506
	Failure	Above Median	-.628	-.209	.214	.326	.790
		Below Median	-.710	.017	-.229	.624	.988
Residual Criteria Test Score	Success	Above Median	507	511	504	506	494
		Below Median	498	468	560	493	498
	Failure	Above Median	499	496	519	500	516
		Below Median	492	520	574	535	544
Fifteen Item Test	Success	Above Median	8.71	8.20	8.00	8.84	9.46
		Below Median	8.17	8.33	10.14	9.90	8.83
	Failure	Above Median	8.09	10.00	8.10	8.70	10.43
		Below Median	8.00	7.75	9.36	8.89	9.89

At the highest ability level (ACE of 142 and higher), the performance of the failure-oriented groups exceeds that of the success-oriented students on the objective measures of performance (Criteria Test Score, Residual Criteria Test Score, and 15-Item Test). With the Criteria Test measures the failure-oriented, low n. Affiliation group is superior to the failure-oriented high n. Affiliation group.

5. DISCUSSION

The negative correlation between debilitating anxiety score and academic performance is expected both from theoretical considerations and earlier empirical results which emphasize the deleterious effects of anxiety upon academic performance (McKeachie, 1951; McKeachie, Pollie, and Speisman, 1955). Likewise, the improved performances of students high in facilitating anxiety tend to support the distinction between the two varieties of anxiety measured by the Alpert-Haber instrument.

Since we found no relationship between the Residual Criteria Test Scores and the Anxiety Scores, or the other personality measures, our results offer some confirmation of the observations of Rust and Ryan (1953) who reported that under- and over-achievers did not differ in emotional adjustment. Our Residual Criteria Test Scores can be interpreted as a measure of Achievement relative to ability and thus serve as an index of over- and under-achievement.

The meager relationships between n. Achievement and n. Affiliation and performance as indicated by the low correlation coefficients seem to suggest that attempts to predict academic performance on these bases are not likely to succeed. This opinion is confirmed by the analyses of the combined effects of motive and ability on performance. We also have evidence that supports the view that the effect of the affiliation motive can best be observed when classroom cues pertaining to the affiliation are taken into consideration (McKeachie, Isaacson, and Milholland, 1964).

The fact that the lower achievement motivation group did better than the higher achievement motivation group in the middle range of ability fails to offer support for the theoretical model proposed by Atkinson in which performance is a joint function of the probability of success (Ps) and the incentive value (Is). On the other hand, the enhanced performance of the middle ability failure-oriented students (Table 4) fails to support the finding of Spielberger (1962), since those for whom anxiety scores were greater than achievement scores showed enhanced rather than diminished performance levels.

A "common sense" interpretation of the high performance of the middle ability high n. Affiliation group and the low performance of the high ability, high n. Affiliation group would imply that high n. Affiliation students in middle ranges of ability see academic performance as an important instrumental activity for fulfillment of present collegiate affiliative goals as well as later post-bacclaureate goals, but that for students high in n. Affiliation also high in ability academic success is practically assured and their attention can be directly centered upon other affiliative acts. Our measures of academic performance reflect performance far below what would be expected only on the basis of ability. While this interpretation is only a plausible hypothesis it could be subjected to empirical confirmation by determining the students' perceptions of the relevancy of grades to long- and short-term goals.

Despite our lack of performance correlates of n. Achievement, we must recognize that this measure of motivation does have correlates in the academic setting. Isaacson (1964), in press, has shown that the selection of concentrate areas is related to n. Achievement scores and in a way predictable from Atkinson's model. People high in the achievement motive do cluster into certain academic areas. Furthermore, the work of McClelland (1961) would suggest that these areas may be related to entrepreneurial professions or careers. The relevance of academic performance to post graduate success in these entrepreneurial careers may vary considerably. It may well be that a failure to obtain academic performance correlates of n. Achievement is due to considerable variability in the perceived instrumentality of academic performance and later achievement goals. Further research is now underway to examine this possibility.

References

Alpert, R. and Haber, R. N. Anxiety in academic achievement situations. J. abnorm. soc. Psychol., 1960, 61, 207-215.

Atkinson, J. W. Thematic-apperception measurement of motives within the context of a theory of motivation. In J. W. Atkinson (Ed.), Motives in Fantasy, Action, and Society. Princeton, D. Van Nostrand Co., 1958.

Atkinson, J. W. and Litwin, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. J. abnorm. soc. Psychol., 1960, 52-63.

Fricke, B. Opinion, Attitude and Interest Survey. Evaluation and Examinations Division, The University of Michigan, 1962.

Isaacson, R. L. Motivation and curricular choices. J. abnorm. soc. Psychol., 1964, in press.

Lowell, E. L. The effect of need for achievement on learning and speed of performance. J. Psychol., 1952, 33, 31-40.

Mandler, G. and Sarason, S. B. A study of anxiety and learning. J. abnorm. soc. Psychol., 1960, 60, 52-63.

McKeachie, W. J. Anxiety in the college classroom. J. educ. Res., 1951, 45, 153-160.

McKeachie, W. J., Isaacson, R. L., and Milholland, J. E. The Affiliation motive, teacher warmth, and academic achievement. 1964, in preparation.

McKeachie, W. J., Pollie, D., and Speisman, J. Relieving anxiety in classroom examinations. J. abnorm. soc. Psychol., 1955, 50, 93-98.

McClelland, D. C. The Achieving Society. Princeton: D. Van Nostrand Co., 1961.

McClelland, D. C., Atkinson, J. W., Clark, R. A., and Lowell, E. L. The Achievement Motive. New York, Appleton-Century-Crofts, 1953.

Morgan, H. H. Measuring achievement motivation with "picture interpretations." J. consult. Psychol., 1953, 17, 289-292.

Rust, R. M. and Ryan, F. J. The relationship of some Rorschach variables to academic behavior. J. Pers., 1953, 21, 441-456.

Ricciuti, H. N. and Sadacca, R. The prediction of academic grades with a projective test of achievement motivation. II. Cross validation of the high school level. Princeton, N.J., Educational Testing Service, 1955.

Shipley, T. E. and Veroff, J. A projective measure of *n*. Affiliation. *J*. *exp*. *Psychol*., 1952, *43*, 349-356.

Smith, C. P. and Feld, Shiela. How to learn the method of concent analysis for *n*. Achievement, *n*. Affiliation, and *n*. Power. Appendix I of Atkinson, J. W. (Ed.), *Motives in Fantasy*, *Action*, *and Society*. Princeton: Van Nostrand Co., 1958.

Spielberger, C. D. The effects of manifest anxiety on academic achievement of college students. *Ment*. *Hyg*., 1963, *46*, 420-426.

Taylor, Janet A. A personality score of manifest anxiety. *J*. *abnorm*. *soc*. *Psychol*., 1953, *48*, 285-290.

Veroff, J., Atkinson, J. W., Feld, Sheila, C., and Gurin, G. The use of thematic apperception to assess motivation in a nationwide interview study. *Psychol*. *Monogr*., 1960, *74*, 499.

V-C. A MODEL FOR CURRICULAR CHOICES*

Robert L. Isaacson

Why does a given college student select one course, or series of courses, rather than others? Why does a student concentrate in physics rather than anthropology? Common sense suggests that a student's choice among the courses and possible concentration areas might be made on the basis of the current cultural emphasis or "glamor value." Other "common sense" interpretations suggests that students elect courses which will best prepare them for their chosen careers. At a somewhat more sophisticated level we might expect that students' collegiate elections reflect "rational choices" based on their perception of their own skills, abilities, and personal values which are related to a choice of careers. Personality theorists might seek to explain course elections on the basis of their role in broad adjustment patterns, although even Freud hesitated to give advice on such important decisions.

It is tempting to accept the proposition that a student's selection of courses is determined by so many variables that attempts at understanding and prediction must be doomed to failure. However, it may be useful to examine the implications of some theoretical assumptions about the selection of courses derived from recent work in the area of motivation in the belief that choice behavior may be more simply understood than first imagined.

In our culture the collegiate educational system almost universally provides opportunities for each young adult to be critically and repeatedly evaluated in relation to others. The college system thereby provides almost an archetype of a competitive situation. Recently, Professors David C. McClelland of Harvard University and John W. Atkinson of The University of Michigan have provided us with a theory of motivation directly concerned with the assessment of motives for success in competition with standards of excellence. They have developed techniques of measuring the need for achievement (_n_. Achievement) in competition with standards of excellence, usually using stories written in responses to pictures.** The pictures used in this method provoke many different kinds of interpretive responses, and these written protocols are objectively scored for the presence and degree of achievement imagery. The scoring technique can be generalized to written responses to ambiguous verbal questions as well as pictures, and the basic theory of motive measurement has been applied to historical funeral orations and even to non-verbal graphic designs.*

---

*The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

**The method used to score imaginative stories for _n_. Achievement is described in McClelland, D. C., Atkinson, J. W., Lowell, E., and Clark, R. A., _The Achievement Motive_, New York, Appleton-Century-Crofts, 1953, and Smith, C. P., and Feld, Shelia, How to learn the method of content analysis for _n_. Achievement, _n_. Affiliation, and _n_. Power, Appendix I of Atkinson, J. W. (Ed.), _Motives in fantasy, action, and society_, Princeton, D. Van Nostrand Co., Inc., 1958.

In opposition to the need for achievement one can postulate a fear of failure in competitive situations. People for whom fear of failure is prominent could be assumed to be motivated to avoid failure. A motive to avoid failure is commonly inferred from scores of subjects on objective "anxiety questionnaires" such as that developed by Mandler and Sarason.** Most important for our present purposes is the fact that it is possible to obtain scores which can be used to represent individual approach or avoidance tendencies in situations in which success or failure is possible.

A point which must be stressed is that motive concepts, like other concepts of psychology, function to predict behavior only when a specific linkage is provided which gives the relation of "motive variables" to other variables and ultimately to performance. Atkinson has proposed that choices among possible courses of action can be predicted on the basis of the total *motivation* toward each of the situational possibilities. *Motivation* is a concept which must be distinguished from the concept, *motive*. For example, the amount of positive *motivation* toward a competitive goal is determined by the product of the latent motive for success (*Ms*), probability of goal attainment (*Ps*), and incentive value of the goal (*Is*).*** The latent motive (*Ms*) is inferred from *n*. Achievement scores obtained from imaginative protocols. Probability of incentive attainment is considered to be the subjective probability (*Ps*) of goal attainment, and the definition of incentive value of the goal is given by (*1-Ps*). This last definition suggests that the amount of satisfaction experienced by an individual reaching a goal is inversely related to the likelihood of goal attainment. In other words the less the subjective likelihood of reaching the goal the more will be the satisfaction if it is obtained.

Failure-avoidance motivation is defined as the product of latent motivation to avoid failure (*Maf*), probability of failure (*Pf*) and the negative incentive value of failure (*If* = -*Ps*). The motive to avoid failure is deter-

---

*See McClelland, D. C. Some social consequences of achievement motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955*, Lincoln, University of Nebraska Press, 1955, and Aronson, E. The need for achievement as measured by graphic expression. In Atkinson, J. W. (Ed.), *op. cit.*

**Mandler, G., and Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1960, *60*, 52-63.

***Atkinson, J. W., and Reitman, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, *53*, 361-366, and Atkinson, J. W., Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In Atkinson, J. W. (Ed.), *op. cit.*

mined by a score from an anxiety questionnaire. The incentive value of failure is defined in such a manner that the greater the likelihood of success the greater will be the dissatisfaction experienced if failure does occur.

These statements can be written:

$$\begin{aligned}\text{Success Motivation} &= Ms \times Ps \times Is, \text{ or}\\ &= Ms \times Ps \times (1-Ps)\\ \text{Failure Motivation} &= Maf \times Pf \times If, \text{ or}\\ &= Maf \times (1-Ps) \times -Ps\end{aligned}$$

In this discussion let us use a value of unity for the motive strengths (_Ms_ and _Maf_) and consider only the relation of the approach or avoidance tendencies toward competitive goals upon curricular choices.

1. SECTION OF COURSES

Let us assume a set of courses: A, B, C, D, and E. Let us also assume these courses differ in their perceived levels of difficulty. Let us assume that each student translates this order of difficulty into a probability of his own successful completion of the course (_Ps_). Students will have different degrees of _motivation_ toward each course and these can be easily determined. Taking as an example a hypothetical student who has the value of _Ps_'s given below, and assuming _Ms = 1_, we can calculate his motivation toward each course in the following manner.

Course:	Ps x	Is = 1-Ps x	Ms x	Success Motivation
A	.1	.9	1 =	.09
B	.3	.7	1 =	.21
C	.5	.5	1 =	.25
D	.7	.3	1 =	.21
E	.9	.1	1 =	.09

Since the motivation toward course C is greater than for any other course, the model predicts that the student in this example would elect course C in preference to all others. The general implication of Atkinson's model, namely that individuals high in motives for success choose tasks of intermediate difficulty, has been supported in laboratory experiments.* Individuals whose behavior is determined by failure-avoidance motives should make quite a different choice in the same situation.

---

*Experimental data supporting the model can be found in Atkinson and Reitman, _op. cit._, Atkinson (Ed.), _op. cit._, and in Atkinson, J. W., Bastian, J. R., Earl, R. W., and Litwin, G. H., The Achievement motive, goal setting, and probability preferences, _J. abnorm. soc. Psychol._, 1960, _60_, 27-36.

When confronted with a choice among the same set of courses, individuals motivated for success and those motivated to avoid failure would have equal but opposite resultant motivation for each course. The motivation for each course for a failure-dominated student is presented below.

Course:	Ps = (1-Ps)	x Vf = (-Ps)	x Maf =	Failure-Avoidance Motivation
A	.9	-.1	1 =	-.09
B	.7	-.3	1 =	-.21
C	.5	-.5	1 =	-.25
D	.3	-.7	1 =	-.21
E	.1	-.9	1 =	-.09

The failure-dominated individual should elect either the most difficult course, A, or the least difficult course, E. He would avoid courses in the middle range of difficulty. Failure-avoidant students will either take courses where the change of failure is remote or hard courses (low Ps) in which the failing is not as painful because so many people will also fail.

An interpretation of this model might assume all of the courses equally acceptable for fulfillment of a college distribution requirement. The theory predicts that the courses believed to be very difficult and very easy should be populated by students motivated predominantly by the desire to avoid failure. The courses in the middle range of difficulty should be elected by students motivated by hopes of success. To make this prediction we must assume that students perceive themselves to be of equal ability and to share common views of the difficulties of the various courses. But, students do differ in abilities and do have different perceptions of their own abilities.

2. DIFFERENCES IN PERCEIVED ABILITY

Each individual has a more or less realistic view of his own academic ability. Just how closely a student's own perception of his abilities corresponds to the perception of his abilities held by others is a separate matter, but within limits the correspondence is probably sufficiently close to assert that students with high abilities will have higher Ps's for all courses than low ability students.

Because of an upward bias of Ps in high ability students and a downward bias of Ps in low ability students, students of high ability primarily motivated toward success should elect courses somewhat above the average difficulty. The low ability students motivated toward success should tend to elect courses somewhat below the average difficulty course.

Individuals motivated by concern over failure should show a different pattern of choices. Due to the decreased likelihood of failure for students of high ability, these high ability students with a predominant motive to avoid failure would tend to avoid the difficult courses altogether. For the low ability students with a considerable amount of the failure-avoidance motive, the more difficult courses represent the least negative choice. In the most difficult course, failure is almost certain and thus there is a minimal negative incentive associated with it.

This analysis would lead us to the remarkable prediction that for all individuals whose predominant motive is to avoid failure, the students of greatest ability would elect the very easiest courses (D and E in our example), whereas the low ability students should elect the most difficult courses (A and B in our example). Said another way, students who pick courses for which they are academically "unfit" should be low in ability but high in the failure-avoidance motive.

3. EFFECT OF PAST EXPERIENCE ON SELECTION OF A SERIES OF COURSES

At this point it would be valuable to consider a student's elections in terms of a possible series of courses. Such situations are common on campus. For example, students must elect areas of courses which fulfill "distribution requirements," sequences in cognates and languages, and areas of concentration. If choices among hypothetical courses, A, B, C, D, and E are considered to be choices among possible sequences of courses then we could predict the behavioral effects of success or failure in early courses in the sequence. Making this change in our units of analysis, we should consider $\underline{Ps}$ or $\underline{Pf}$ as the probabilities of success or failure in the sequence as a whole.

The successful performance in the initial course in a series should have the effects of raising the $\underline{Ps}$ for the entire series. If a success-oriented student originally believes his likelihood of completing a sequence C is .5, and he does very well in the first course, his estimate of the likelihood of success in sequence C may be raised to .6, or more. With this increase in subjective probability there would be concomitant decrease in total motivation. A student under these conditions might switch to another and seemingly more difficult sequence which is now perceived as offering a $\underline{Ps}$ closer to .5. Poor performance in the initial course would tend to alter the $\underline{Ps}$ downward and make the selection of a less difficult sequence more likely. These possible changes in elections of the success-motivated students are those we usually call "reasonable." The behavior predicted for students motivated by concern over avoiding failure once again appears paradoxical.

When the failure-motivated student who has elected the most difficult course sequence succeeds in the initial course, he should, according to the theory, drop the sequence and choose the easiest available sequence to fulfill the requirement. This result is predicted on the basis of the change in the probability of failure (*Pf*) for the entire sequence. As a function of success, the *Pf* for the sequence becomes smaller. If failure in the sequence was thought to be almost certain (*Pf* ~.9) the initial success may change it to something much more likely (*Pf* ~.7). This would produce an increase in the total motivation *to avoid* the particular sequence. The easiest sequence might appear as easy as it had been before and if its original *Pf* had been near .1 then it would now be the least undesirable choice. In this way the model would predict a shift from the hardest to the easiest sequence, given success in the first course. In a similar manner the model would predict failure in the first course of the easiest sequence should elicit a shift to election of the most difficult series of course in the failure-oriented students.

4. THE EFFECT OF MOTIVES ON Ps AND Pf

In Atkinson's motivational model differences in the strength of motives are assumed to have two separate effects: (1) To alter the value of the number used to represent, *motive*, in the basic formula: motivation = incentive x probability x *motive*; and (2) to bias the individual's estimate of *Ps*.

Thus far, the motive term in the formula has been set at unity. To have any choice behavior the latent motive for success or to avoid failure must be greater than zero. Differences in motive levels would be represented by insertion of different numbers for the value of latent motive. This would produce the effect of accentuating the predicted pattern of results but the pattern of choices would remain unchanged.

It could be assumed that a person highly motivated for success has generally higher *Ps*'s than those persons less motivated toward success. This would result in effects similar to those predicted for high ability students. In the other extremes people strongly motivated away from failure might tend to see *Pf* as greater (*Ps* lower) than people less strongly motivated. This would result in choices similar to those made by low ability, failure-motivated students.

It should be noted that the model which we have been discussing may be complicated but it does not include complex interactional processes found in more elaborate models of personality. Given the measures of motivation for an individual and his estimates of success in a choice among courses or sets of courses, predictions can be generated which are susceptible to direct empirical test. As we have seen, predictions about changes in election can be generated by simple assumptions as to the effects upon changes in an individual's subjective probabilities of success or failure.

5. TESTING THE MODEL

Tests of many aspects of the basic motivational model have been made by numerous experimenters.* Recently it has been possible to test the model insofar as it applies to curricular choices. The areas of concentration of a number of honors students and superior non-honors students at The University of Michigan for whom motivational scores were available were determined. As freshmen, these students wrote imaginative stories to the French Test of Insight** which were scored for *n*. Achievement. Also as freshmen, they answered the Mandler-Sarason Test Anxiety Questionnaire.*** Scores from these tests were used to estimate the extent to which an individual was predominantly motivated for success or to avoid failure.

The students were divided into two groups of men or women on the basis of their *n*. Achievement and Test Anxiety scores.† One group represented an extreme group which was almost exclusively achievement oriented while the other group was almost exclusively failure-avoidance oriented. The areas of concentration were classified into three categories representing relatively easy (high *Ps*), hard (low *Ps*), or intermediate choices.†† If one assumes the difficulty classification to be close to the subjective probabilities of academic success of the students then the theory would predict the achievement-oriented men to elect the intermediate concentrate areas whereas the failure-avoidant

---

*For substantiation of this statement one need only look at the substantial number of experimental contributions to be found in *Motives in fantasy, action, and society*, Atkinson, J. W. (Ed.), *op. cit*.

**French, Elizabeth G. Development of a measure of complex motivation. In Atkinson, J. W. (Ed.), *op. cit*.

***Mandler, G., and Sarason, S. B., *op. cit*.

†The data to be described are reported more completely in Isaacson, R. L. The relation between *n*. Achievement, Test Anxiety, and Curricular Choices, 1964 (in press, *J. abnorm. soc. Psychol*.).

††The basis for classification of concentration areas into ordered degrees of difficulty came from estimates of difficulty of specific courses found in the programs. Information was obtained from recent reports of the Evaluation and Examinations Division, The University of Michigan. From these reports it was possible to determine the average grades and average brightness of students found in certain courses. It was assumed that the ratio of average grades to average brightness represent an index of the difficulty of the course. Departmental concentrate programs which had difficult, intermediate, or easy courses (determined in this fashion) were assumed to have concentrate programs of similar difficulty. In addition the author obtained estimates from introductory psychology students as to how well they thought they could do scholastically in the programs which make up the three levels of difficulty groups.

The students' estimates of how well they could do in various concentration programs matched the prior categorization on the basis of the ratios of average brightness to average grades.

men would be found in either the easiest or most difficult areas.*

The curricular choices of the students highly motivated for success (upper quartile) and highly motivated to avoid failure (lower quartile) are presented in Figure 1.** As can be seen in this figure, the achievement-oriented men in the honors and non-honors groups elect the concentration areas considered to be in the middle range of difficulty. The choices of the failure-avoidant men are found in the extreme Ps areas. From these data it is not possible to evaluate the finer detail of the model's predictions. The students on the non-honors groups were superior students whose grade point averages, American College Entrance Examination (total), and motive scores did not differ from students in the honors program. As far as it goes, however, the analysis would seem to be encouraging for the model discussed in this paper.

The model of motivation which I have sought to extend to choices in a collegiate setting is a child of laboratory investigations. Yet it was developed with the hope that a theory based on experimental foundations would be general in scope. Whether or not this theory is found to be sufficiently general and fruitful in the prediction and explanation of behavior, its extension to academic behavior represents a new direction for combined research efforts of men in psychology and education. As noted by Spence a few years ago the time is ripe for the testing and subsequent use of theoretical concepts coming from the laboratories in the crucible of the instructional situation.***

---

*Only the choices of the high n. Achievement (upper quartile) and high failure-avoidance men (lower quartile) are reported. Achievement scores obtained from women have tended to present a rather separate, and unsolved, problem. See pp. 172-81 in McClelland, Atkinson, Lowell, and Clark, op. cit.

**The data to be described are reported more completely in Isaacson, R. L. The relation between n. Achievement, Test Anxiety, and Curricular Choices, 1964 (in press, J. abnorm. soc. Psychol.).

***Spence, K. W. The relation of learning theory to the technology of education. Harvard Educ. Rev., 1939, 29, 84-95.

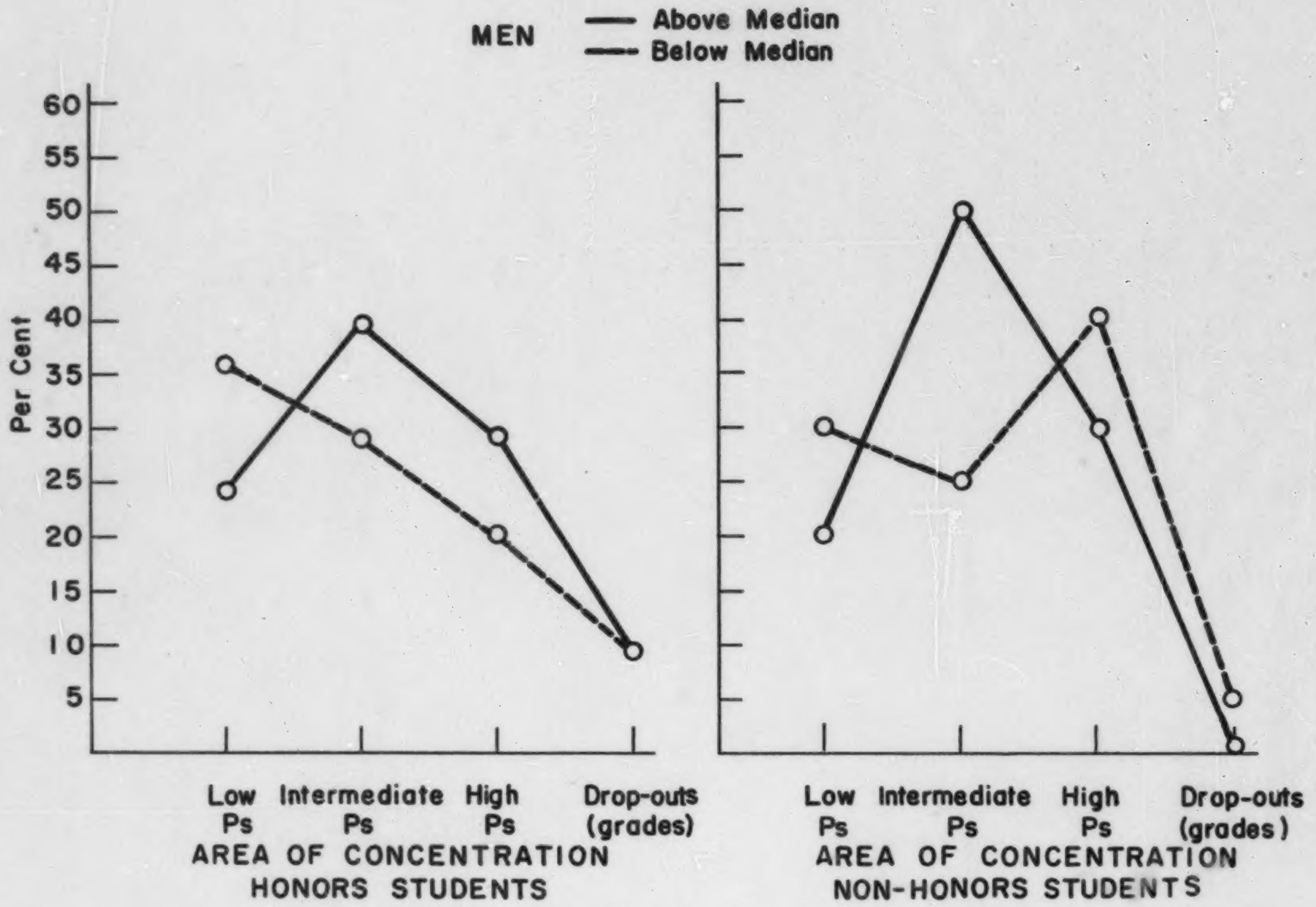

Figure 1. The percent of men above the upper quartile of n. Achievement (upper quartile) and below the lower quartile of Test Anxiety (lower quartile) men in the honors program (left) and superior nonhonors men (right) electing concentration areas of different subjective difficulty levels. Also shown are the percentages of each motive category definitely known to have dropped out of college because of poor grades. The figure is based on data reported by Isaacson.

V-D. THE RELATION BETWEEN n. ACHIEVEMENT, TEST ANXIETY AND CURRICULAR CHOICES*

Robert L. Isaacson

Abstract

A model of motivation proposed by Atkinson was considered appropriate for the generation of hypotheses about the curricular choices of college students. The selection of areas of concentration by honors and superior nonhonors students, tested as freshmen, were obtained when the students were seniors. Their choices were related to their n. Achievement scores and Test Anxiety scores. The general predictions of the model were borne out for men but not for the women, although this was not unexpected. In regard to the choices of women it was noted that women in the honors program tended to obtain teaching certificates if they had high Test Anxiety scores.

Atkinson has proposed a model of choice behavior in which latent motives, probability of goal attainment, and the incentive value of the goal are differentiated at a theoretical and empirical level (Atkinson, 1958). In this model motivation toward a goal is a product of these three terms. In a situation where several courses of action are possible the choice among the possibilities is made on the basis of the motivation toward or away from each possible choice. "Latent motives" are measured by projective tests or objective self-report questionnaires. The "probability of goal attainment" refers to the subjective probability that a given course of action will result in the attainment of states relevant to broad motivational goals (i.e., success in competition with a standard of excellence or the avoidance of failure experiences), and is inferred most often from the structure of the experimental situation. The "incentive value of the goal" is defined as 1—(the probability of goal attainment). Tests of his model have been made in laboratory settings (see Atkinson and Litwin, 1960, also Atkinson, Bastian, Earl, and Litwin, 1960).

There is no reason why tests of the model could not be made in a more

---

*This research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

"natural setting." Specifically, an examination was made of the patterns of choices of concentration areas of senior students for whom n. Achievement (McClelland, Atkinson, Lowell and Clark, 1953) and Test Anxiety (Mandler and Sarason, 1952) scores were obtained when they were freshmen. The difficulty level of the different areas of concentration were estimated to provide a basis of prediction of curricular choices for the different motivational groups.

The model proposed by Atkinson predicts that people high in n. achievement should prefer choices where the probability of goal attainment (Ps) is of intermediate value. This prediction stems from the fact that the resultant motivation for intermediate choices should be greatest at these points since it is a product of the probability of success (Ps) and the incentive value of the goal (1-Ps). People high in n. achievement should not make choices which are either too easy or too difficult because of the low resultant motivation associated with these extremes.

If one assumes that the Test Anxiety score represents a latent motive to avoid failure (Atkinson and Litwin, 1960), then individuals with high Test Anxiety scores should elect extreme choices. They should avoid the middle range of difficulty which is associated with the greatest resultant negative motivation. Accordingly, they should pick the most difficult or least difficult tasks. This paper examines the curricular choices of some selected graduating seniors at The University of Michigan to determine if such patterns of choice could be found. Specifically, students who have greater n. Achievement scores than Test Anxiety scores whould show a greater frequency of election of concentration areas of intermediate difficulty and students who have greater Test Anxiety Scores than n. Achievement scores should be found to elect the most and least difficult areas of concentration.

1. METHOD

Subjects

During the middle of the fall semester 1958-59, 307 freshmen, enrolled in sections of English Composition designed for superior students, were given the Test of Insight developed by Elizabeth French (1958), and the Test Anxiety Questionnaire developed by Mandler and Sarason* (1952). Written responses to ten verbal descriptions of behavior were scores for n. achievement following the procedure described by McClelland, et al. (1953) by experienced judges who had demonstrated scoring reliabilities over .90 on earlier materials.** Form

---

*The author is indebted to Dr. John W. Atkinson for use of the test data reported in this article.

**n. Achievement scoring was done by Mr. Willard Larkin (men), and Mrs. Nancy Tucker (women).

II of the French Insight test was used with different lead-ins for men and women. Approximately half the students tested were in the Freshman Honors Programs; 130 of them were men, and 177 were women. In 1962 it was possible to determine the fields of concentration for 108 men and 109 women of the original samples. In addition it was possible to determine the number of cases in which a student dropped out of The University of Michigan for reasons of poor academic performance.

For the purpose of this study all fields of concentration elected by students were collapsed into three categories: Difficult (low Ps), intermediate difficulty (intermediate Ps), and easy (high Ps).* For the women a separate category was established for those who earned a teaching certificate.

The classification of concentration areas into ordered degrees of difficulty was based on estimates of difficulty of specific courses in academic programs. Information was obtained from recent reports of the Evaluation and Examinations Division. From these reports it was possible to determine the average grades and the average brightness of the students found in certain courses as measured by tests in the freshmen batteries. It was assumed that the ratio of average grades to average brightness represented an index of the difficulty of the course. Departmental concentrate programs which had difficult, intermediate, or easy courses, determined in this fashion, were assumed to have concentrate programs of comparable difficulty. In addition, the author obtained estimates from introductory psychology students as to how well they thought they could do scholastically in the programs which make up the three levels of difficulty groups. Students who had high grade point averages responded without discrimination among three groups of departmental programs. For example, the brightest students responded that they could do well in any area. Students with average and low grade point averages, however, responded in ways which indicated that the previously established order of difficulty was appropriate. These students estimated they could do best in the courses labelled "easy," and least well in courses labelled "difficult." The "intermediate" courses fell in the middle.

Only those students for whom n. Achievement scores and Test Anxiety scores were available and whose scholarly careers could be unequivocally determined were considered in the analysis. Despite careful checking, it was impossible to ascertain the academic progress of all those that took the tests as freshmen. Data will be presented for those who were known to have dropped out because of poor academic performances but not for those for whom the reason for withdrawal from the University was in any way uncertain.

---

*The names of the concentration programs which were labelled "difficult" and "intermediate" and "easy" will be omitted to save pointless controversy. However, it might be added that they correspond fairly well to the student stereotypes of courses found in most colleges.

The distribution of n. Achievement scores and Test Anxiety scores were calculated separately for men and women. No differences were found for either sex in the mean of the raw n. Achievement or Test Anxiety scores of those in the honors program as compared with the superior nonhonors students. The motivational scores of each student were transformed into standard scores (z scores). Two new distributions (one for each sex) were obtained in which the differences between the z score transformations of the n. Achievement scores and the z score transformations of the Test Anxiety scores were ordered. These new distributions were used as the bases for all further analyses. Individuals which have a positive difference score in these distributions represent individuals, for whom the positive motive toward achievement is greater than the negative, failure avoidance, motivation reflected by the Test Anxiety score. It will be assumed, along with Atkinson and Litwin (1960), that Test Anxiety scores reflect a motive to avoid failure.

2. RESULTS AND DISCUSSION

Figure 1 presents the distribution of men in the honors and nonhonors groups falling above or below the median score of the distribution of the differences between the transformed n. Achievement and Test Anxiety scores. The men with scores above the median will be described as the achievement group whereas those below the median will be described as the failure-avoidance group. The achievement men in the nonhonors group clearly evidence the predicted selection of concentration areas of intermediate probability of success (intermediate Ps). The same trend can be observed in the honors group although the effect is not as pronounced.

Since those falling in the middle range of the distribution of the differences between the transformed Achievement and Test Anxiety scores are likely to have mixed patterns of motives, a purer measure of achievement and failure-avoidance motives would be obtained by consideration of only those falling above the upper and below the lower quartiles of the distributions.

Figure 2 presents the concentrate areas for those purer representatives of achievement and failure-avoidance motives. Patterns of curricular choices similar to those shown in Figure 1 are obtained but the differences between the groups are accentuated, as should be the case.

Similar plots of the choices of concentration areas for women are presented in Figures 3 and 4. Figure 3 presents a comparison of the curricular choices of the achievement and failure-avoidance groups based on whether or not the person was above or below the median of the distribution of differences between the transformed Achievement or Test Anxiety scores.

Figure 4 compares the choices of women falling above the upper quartile and below the lower quartile of the same distribution.

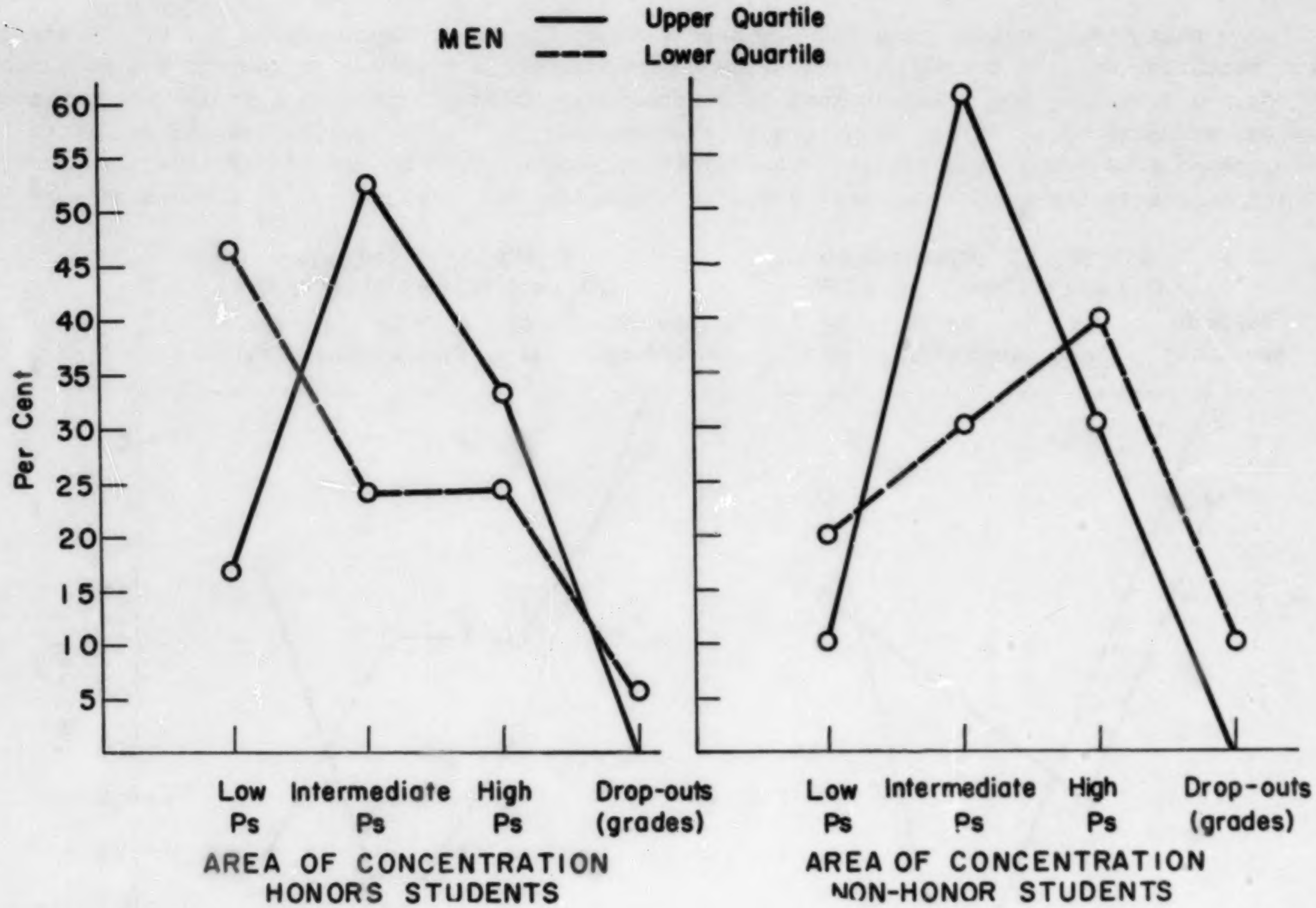

Figure 1. The percentage of men above (achievement group) and below (failure-avoidance group) the median of the distribution of the differences between the transformed n. Achievement and Test Anxiety scores concentrating in low, intermediate, and high Ps areas. Also shown is the percent of each motivational group withdrawing from college because of poor grades. The left-hand portion of the figure describes the choices of men in the honors program while the right-hand portion describes the choices of men not in the honors program. N of honors students = 68; N of nonhonors students = 40.

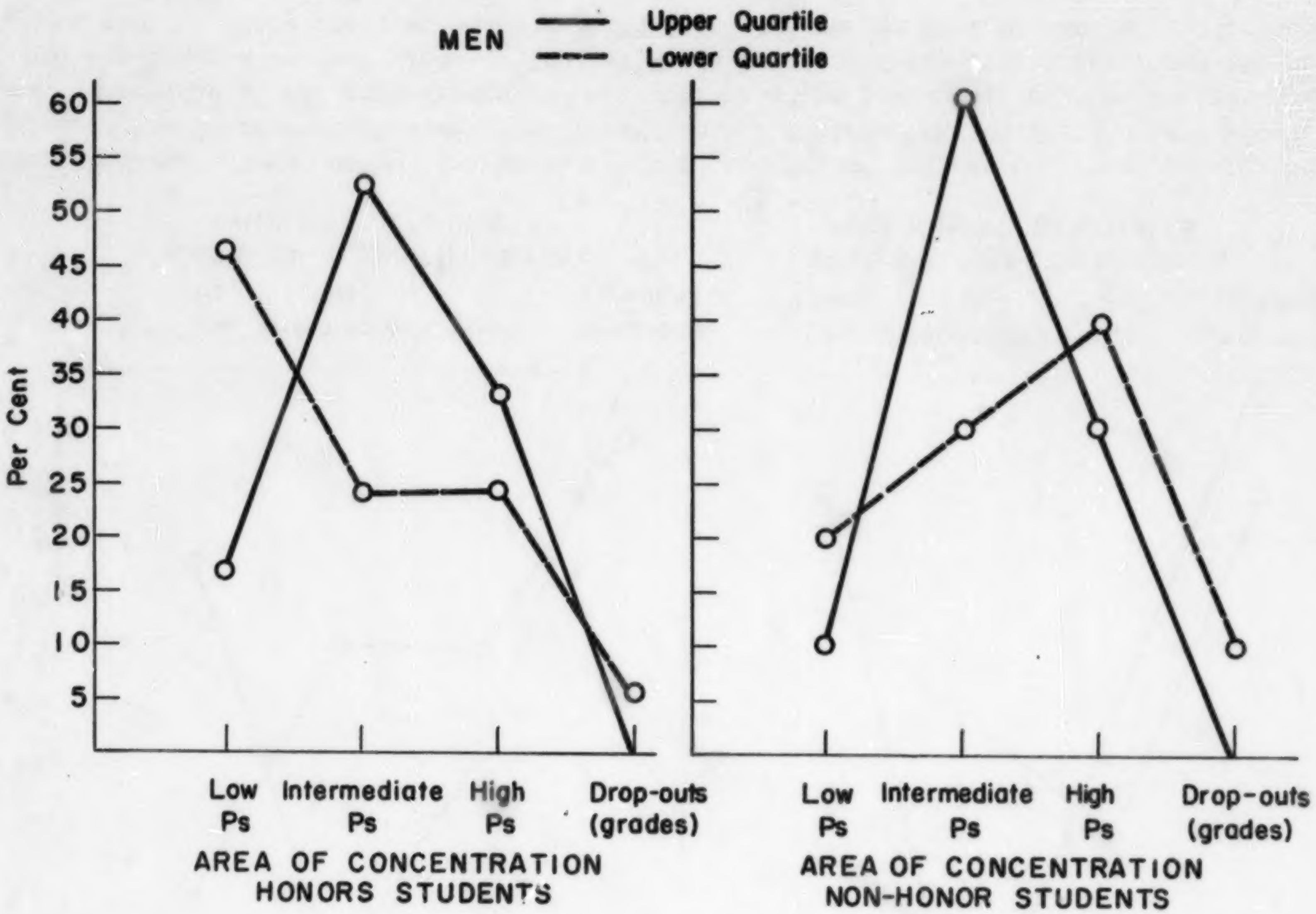

Figure 2. The percentage of men above in the upper quartile (achievement group) and below the lower quartile (failure-avoidance group) of the distribution of differences between the transformed n. Achievement and Test Anxiety scores concentrating in low, intermediate, and high Ps areas. Also shown is the percent of each motivational group withdrawing from college because of poor grades. The left-hand portion of the figure describes the choices of men in the honors program while the right-hand portion describes the choice of men not in the honors program. N of honors students = 35; N of nonhonors students = 20.

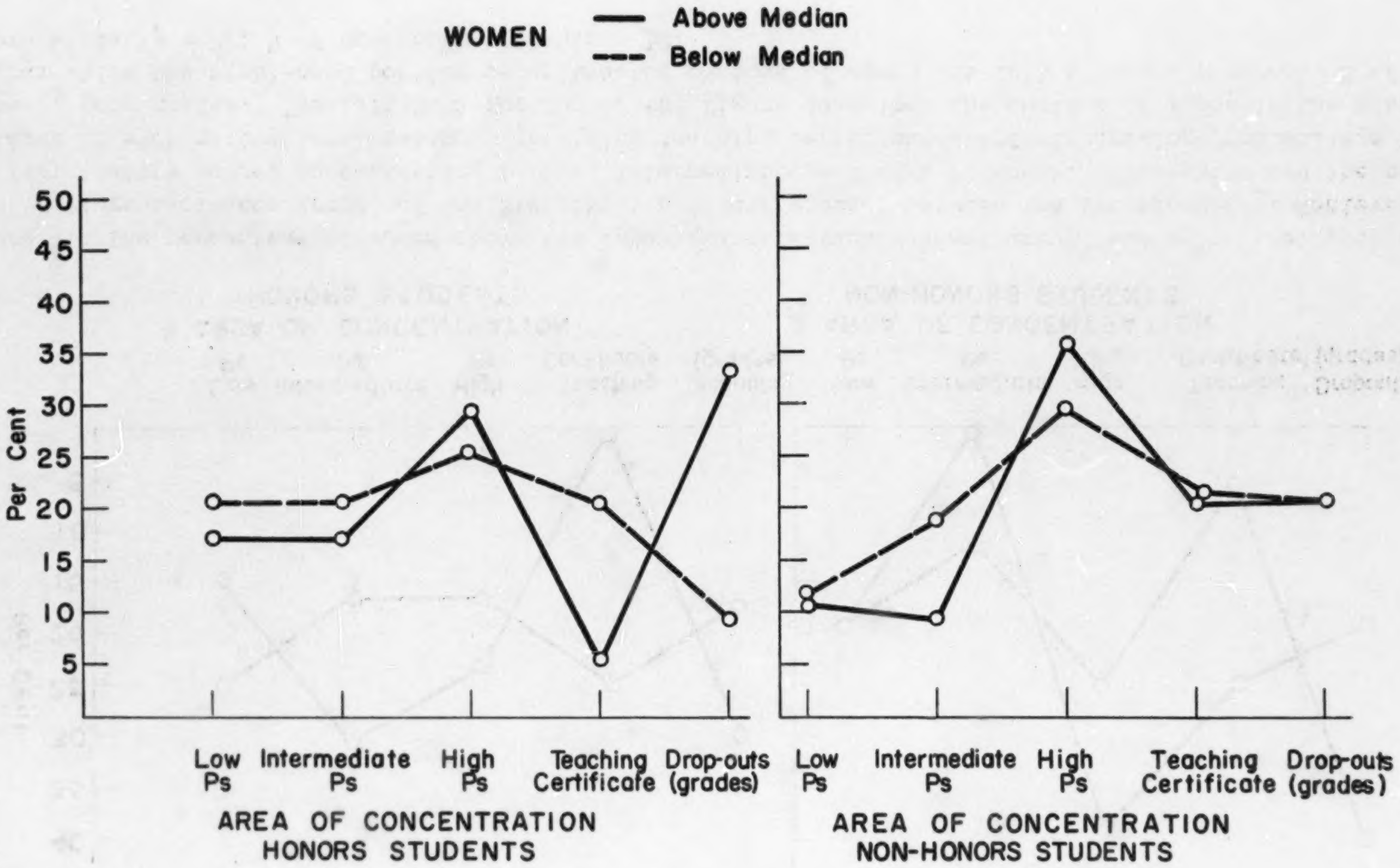

Figure 3. The percentage of women above (achievement group) and below (failure-avoidance group) the median of the distribution of the differences between the transformed n. Achievement and Test Anxiety scores concentrating in low, intermediate, and high Ps areas. Also shown are the percentages of each motivational group obtaining teaching certificates and withdrawing from college because of poor grades. The left-hand portion of the figure describes the choices of women in the honors program while the right-hand portion describes the choices of women not in the honors program. N of honors students = 42; N of nonhonors students = 67.

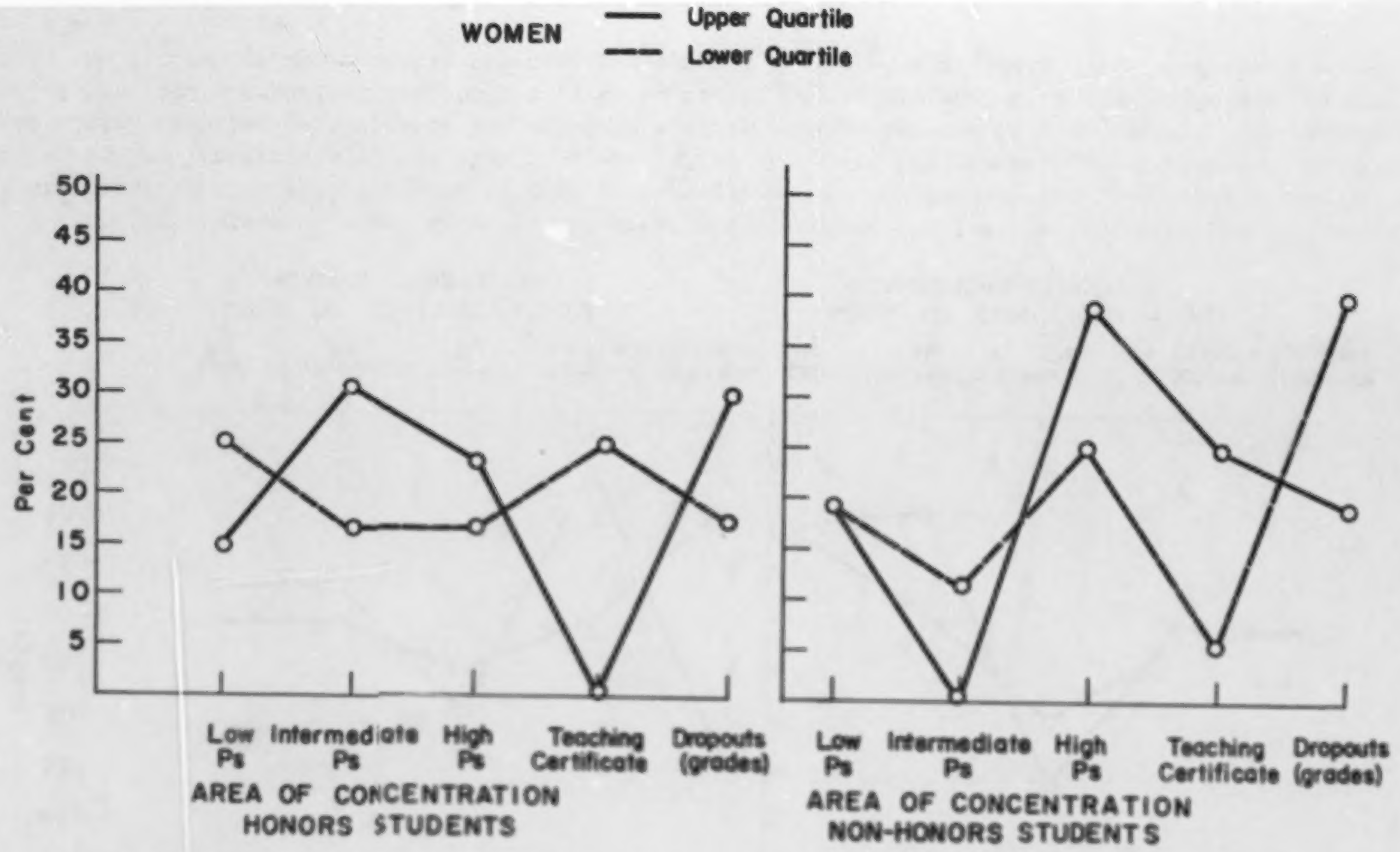

Figure 4. The percentage of women above the upper quartile (achievement group) and below the lower quartile (failure-avoidance group) of the distribution of differences between the transformed n. Achievement and Test Anxiety scores concentrating in low, intermediate, and high Ps areas. Also shown are the percentages of each motivational category obtaining teaching certificates and withdrawing from college because of poor grades. The left-hand portion of the figure describes the choices of women in the honors program while the right-hand portion describes the choices of women not in the honors program. N of honors students = 22; N of nonhonors students = 32.

The all over relationship between achievement and failure-avoidance motives found in the data obtained from the men does not appear in the data from the women. This result is not unexpected because the projective measures used to assess *n*. achievement have not been generally successful. The *n*. achievement measure for women has been thought to constitute a rather separate problem (see McClelland *et al*., 1953, pp. 172-181). On the other hand women in the honors program who were the most representative of achievement motivation (Figure 4) showed a tendency to choose concentrate areas of intermediate probabilities of success. Women of a similar motivational pattern in the nonhonors group, however, showed a preference for the concentrate areas with the greatest Ps.

There seems to be a rather striking difference between the curricular choices women in the honors program and those not in the honors program. The high failure-avoidance women in the honors program tend to take programs leading to teaching certificates whereas this difference tends to wash out or actually be reversed in the nonhonors group, at least for women in the honors program the acquisition of teaching certificates seem closely related to high scores on the Test Anxiety questionnaire.

While the men's choices of concentration areas seems to support the general model proposed by Atkinson, there are other theoretical predictions which would offer convincing evidence if supported. Atkinson's model would predict that men motivated primarily by failure-avoidance motives who select low Ps concentration areas should be ones who have had experiences of failure in either high *Ps* or low *Ps* areas. Success obtained at the highest *Ps* choices would make these choices seem to be easier than they had supposed and thereby make this alternative even more attractive. Students electing such areas of concentration would continue in them. Success in the lowest Ps areas would tend to make success in them more likely and reduce the attractiveness of the courses.

On the other hand failure-avoidance motivated men who select the concentrate areas with the greatest probabilities of success should have histories of success at these easiest areas or successes in the concentrate programs with the least probabilities of success. Unfortunately, it was not possible to test these hypotheses adequately with the current data. The students in both the honors and special English sections were very able and there was no difference in either total grade point average or ACE scores (total) between the failure-avoidance men majoring in the high or low Ps areas.

In general the results obtained from men, at least, would suggest that the model proposed by Atkinson can be brought to bear upon problems which ordinarily are thought to be explicable only in terms of more elaborate theoretical models. The few concepts used in the model can be easily reduced to empirical observation and therefore invite future use in studies of decisions in the academic setting.

References

Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, Van Nostrand, 1958.

Atkinson, J. W., Bastian, J. R., Earl, R. W., and Litwin, G. H. The Achievement motive, goal setting, and probability preferences. *J. abnorm. soc. Psychol.*, 1960, *60*, 27-36.

Atkinson, J. W. and Litwin, G. H. Achievement motive and test anxiety conceived as a motive to approach success and motive to avoid failure. *J. abnorm. soc. Psychol.*, 1960, *60*, 52-63.

French, Elizabeth G. Development of a measure of complex motivation. In Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, Van Nostrand, 1958.

Mandler, G. and Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, *16*, 115-118.

McClelland, D. C., Atkinson, J. W., Lowell, E., and Clark, R. A. *The Achievement Motive*. New York, Appleton-Century-Crofts, 1953.

Student Affiliation Motives, Teacher Warmth, and Academic Achievement[1]

W. J. McKeachie

This is one of a series of studies concerned with interactions between student motives and cues from instructors that affect achievement in college classes. The theoretical context for these studies is described in the first report of the series (McKeachie, 1961).

> "Students enter college with a number of important, relatively stable motives learned at home, in their previous educational experiences, and through other experiences. There are not only individual differences in these motives but also differences in the relative strength of different motives within individuals. Students also possess skills and abilities useful in achieving goals in college. Some of these, such as verbal ability, are measured by the intelligence tests given for admission: others like note-taking skill, study habits, ability to gain the instructor's sympathy, and ability to make friends, are not measured in our entrance battery.

Unfortunately for the sort of research described in this paper, there is a motivational factor of major importance common to all classes and students - grades. Whatever his motives for attending college, a student is motivated for grades. At the very least he must maintain an average high enough to remain in college, and in most cases his access to vocational or other goals depends to some extent upon achievement of higher grades. This is probably particularly true at Michigan where the student body is highly selected on the basis of high school grades and the social climate emphasizes academic achievement. Because most students are highly motivated for grades, the differential effects of differing teaching methods in eliciting such motives as need for achievement, need for affiliation, and need for power are difficult to detect,

---

[1] The earlier data reported in this study were collected under a grant from the Fund for Advancement of Education.

Nevertheless, if everyone were motivated equally the correlation between ability measures and grades should be much higher than it is. Even the best combinations of measures of ability leave a good deal of variance in criterion measures unaccounted for - enough at least to leave hope that motivational measures may contribute significantly.

The total college environment has much to do with activating motives and skills. As Pace and Stern (1958) have shown, colleges differ markedly in their students' perceptions of what are important cues in the college environment.

At the same time that the individual's motives are being activated, he is choosing a strategy,[2] such as maximizing satisfaction or minimizing regret. Since the student is involved in several courses as well as extracurricular, social, and other activities, his strategy must involve assessment of the potential rewards and punishments and their probabilities of attainment in all of these situations. Whatever his strategy, its choice also involves an assessment of his habits and skills in relation both to the possible satisfactions in the situation and to the types of behavior necessary to achieve these satisfactions. From these dual assessments come an estimate of the amount and kind of satisfaction and the probability of attaining it in each class. This combination of expectancies of amount of affect and of probability of affect is what will be called "motivation"[3] herein.

The individual's habits and skills thus enter into motivation in determining subjective probability. But they also enter directly into predictions of achievement, for if we are to predict a grade we must assess the instrumental habits actually used by the student to achieve his goals as well as his real skill in their use.

---

[2] The words "choice" and "strategy" may imply a much greater degree of consciousness and rationality than intended. Clearly much of the process is unconscious and intuitive.

[3] In our usage, "motive" refers to the stable tendency to receive satisfactions from attaining a class of goals, and "motivation" refers to the strength of the tendency to strive for a particular goal at a particular time.

> An example may illustrate this. Suppose that a student has a strong achievement motive in a classroom in which the teacher provides cues to standards of achievement and incentives for achievement. This combination arouses expectancies of achievement satisfaction for the student, but the strength of his motivation will also be influenced by his estimate of the chances of achieving the standards set by the instructor. Atkinson (1958) has suggested that motivation for achievement will be highest when subjective probability is moderate, i.e., when the uncertainty of the outcome is at its maximum.[4] Two students with strong achievement motives may select the same instrumental habits to achieve their goals; e.g., they may both do their homework, but if they differ in subjective probability of success, their motivation will differ and we may expect differences in attentiveness and persistence and ultimately in achievement. Similarly two students may both have strong achievement motives and both have the same expectancies of success, yet they may differ in the instrumental habits they choose to attain success or in the skill with which they actually perform the instrumental responses chosen, and these differences may be reflected in their ultimate achievement."

Our hypothesis was:

> The academic achievement of students high in affiliation motivation will be relatively higher in classes high in affiliative cues than in classes with few affiliative cues, while the achievement of students low in affiliation motivation will be relatively lower in classes high in affiliative cues than in classes with few affiliative cues.

The thinking leading to the hypothesis stated above began by assuming that the affiliation motive is an important dimension of individual differences in college students. Classes differ in the extent to which they provide cues for this motive. Some instructors are warm, friendly, and personally interested in each student: other instructors are subject or self-oriented and elicit few expectations of affiliative satisfactions. Assuming that the instructor arousing affiliative motivation also provides affiliative satisfactions, we would expect students high in need for affiliation to show greater satisfaction in such classes than in those taught with less affiliative cues.

---

[4] Although Atkinson has made his case only for achievement motivation, it is not improbable that a similar interaction might hold for other motives.

The gap in this theory is that even when a student wants to win affiliative satisfaction from the teacher it is not certain that he will perceive academic achievement as being instrumental to such satisfactions. Quite conceivably he might perceive apple-polishing, asking questions or other behaviors as instrumental responses with a greater probability of arriving at his goal of affiliative satisfaction. These instrumentalities were not investigated in our studies.

Procedures

Sample

To test our hypothesis we needed a sample of students differing in motivation and a sample of instructors differing in the cues emitted. Since we were primarily concerned with relative achievement within a class, it would have been possible to use a number of sections of different courses. There were, however, some advantages to having several teachers trying to achieve the same goals, since this permitted comparing over-all effects of differing methods. At the same time, however, there were clear advantages for generalization if conclusions could be based upon more than one course. Consequently, for our first study we chose three multi-section courses representing quite different types of content and objectives. They were second-year French, freshman mathematics, and general psychology. Thirty-one instructors participated - all of the experienced teachers in these courses.

Within these instructors' sections all students were invited to take a group of tests designed to measure our independent variables. Of the 825 students in the three courses, 609 provided complete and usable responses on all tests. There were 208 students in French, 188 in mathematics, and 213 in psychology. Two hundred fifty-four were men; 355 were women.

In the second study 16 introductory psychology teachers participated. There were 406 women and 348 men enrolled in the introductory course, of whom 737 took the measures of achievement.

Measures

The TAT method developed by Shipley and Veroff (1952) was used to measure student need for affiliation. The pictures were those used in the national study of motivation by Veroff, Atkinson, Feld, and Gurin (1960). The reliability of our trained scorers, as measured by correlations with experts, ranged from .82 to .89.

Affiliation Cues

In the first study each student completed a twelve-item questionnaire about characteristics of the classroom. This questionnaire contained the three items with the highest factor loadings on each of four factors emerging from a factor analysis of a longer questionnaire. The four factors had to do with achievement cues, friendliness, organization of the content of the course, and organization of testing and grading.

The items for affiliation cues (friendliness) were as follows:

> The instructor takes a personal interest in students.
> Students in the class were friendly.
> The instructor calls students by name.

The mean of students' ratings on a 1-5 scale was used as the index of the level of affiliation cues in a class.

In the second study, classroom observers rated the overall level of affiliation cues.

Outcome Measures

Two types of outcome measures were used. Course grades were used in all courses and achievement tests were given to the psychology students. In the first study psychology students had as part of their final examination 15 multiple-choice items and one essay question. All essay tests were scored blind by a single experienced instructor who was not involved in the experiment. The second study used the same 15 item objective test and the "Criteria Test" (Milholland, 1964) as part of the final course examination.

Results and Discussion

The experimental hypothesis was supported in the first study by a significant interaction in the expected direction among affiliation cues, motivation and grades for all students in psychology and for men in mathematics and psychology taken together. (See Table 1) Similar results were obtained for men in the second study. (See Table 2) There was a slight tendency for high n Aff women in the second study to perform as predicted, but the grades for low n Aff women were contrary to the hypothesis.

The results lend satisfying support to our hypothesis. But with grades as the principal criterion, it is possible that the students with strong affiliation motivation attained their highest grades in high

Table 1

The Interaction of Affiliation Cues, the Affiliation Motive and Grades (First Study)

Mathematics (men)

N Aff	Aff Cues	A + B	C + D	Total
H	H	19	12	31
	L	11	18	29
L	H	22	26	48
	L	10	27	37
		62	83	145

Psychology (men)

N Aff	Aff Cues	A + B	C + D	Total
H	H	8	8	16
	L	8	13	21
L	H	9	19	28
	L	11	9	20
		36	49	85

Psychology (women)

N Aff	Aff Cues	A + B	C + D	Total
H	H	19	17	36
	L	17	17	34
L	H	13	23	36
	L	22	20	42
		71	77	148

	Chi square	df	P
_n_ Aff X Aff cues X Grades (Men)	5.2	1	.05
_n_ Aff X Aff cues X Grades (Psychology)	8.5	1	.01

Table 2

The Interaction of n Affiliation and Teacher Affiliation Cues with Grades

(Second Study)

MEN

Student n Aff	Aff Cues	A and B	C, D, E	Total
Hi	Hi	36	21	57
	Lo	26	33	59
Mid	Hi	22	27	49
	Lo	21	25	46
Lo	Hi	20	25	45
	Lo	31	25	56
		156	156	312

Variables	Chi square		p
n Aff x Aff cues	1,034	2	-
n Aff x Grade	1,414	2	-
Aff cue x Grade	.321	1	-
n Aff x Aff cues x Grade	5,473	2	.10

WOMEN

Hi	Hi	43	44	87
	Lo	50	58	108
Lo	Hi	47	38	85
	Lo	40	53	93
		180	193	373

cue classes simply by making a more favorable impression upon the instructor.

Fortunately we have one check upon this in the scores on the common portion of the final examination in psychology. As Table 3 indicates, the pattern of results on the objective examination reveals no interaction. On the essay test only the women tend to perform as predicted. On the Criteria test used in the second study however, high n Aff men tended to do well in high affiliation cue sections, but the performance of women was opposite to what was expected. On the test of knowledge we find a significant interaction with both sexes performing as predicted. (See Table 4)

Discussion

Why did men high in n Aff achieve better grades for friendly teachers? Our guess would be that these students saw academic achievement as a means of gaining the approval and personal interest of the teacher. But this guess should be checked, and we anticipate questioning such students about their perceptions of avenues to affiliative satisfactions.

The results reported here may be related to those of recent studies of public school teachers by Cogan (1958) and by Ryans (1960), which indicate that a dimension characterized by friendly, integrative, affiliative, nurturant, behavior is important in public school teaching. Teachers high in this dimension produced more student self-initiated work than teachers low in the dimension. Although teacher warmth appears to be an important dimension of teaching, our results indicate that its effect is not uniform. The "warm" teacher (high affiliation cues) is not always effective. His effectiveness depends upon characteristics of his students.

Summary

The hypothesis that affiliation cues would interact with n Affiliation is determining achievement in a college class was tested in introductory courses in French, mathematics, and psychology. Both in the original study and in a replication study men high in n affiliation made better grades in classes characterized by a high level of affiliation cues whereas low n Aff men did better in classes low in affiliation cues.

Table 3

The Interaction of n Affiliation, Teacher Affiliation Cues and Performance on Final Course Examination in Psychology

(First Study)

Objective Test of Knowledge
Men

n Aff	Aff Cues	Hi	Lo	
H	H	7	7	14
	L	11	10	21
L	H	14	14	28
	L	11	9	20
		43	40	83

Objective Test

Women

n Aff	Aff Cues	Hi	Lo	
H	H	12	24	36
	L	18	16	34
L	H	15	19	34
	L	21	20	41
		66	79	145

Essay Test Scores

Men

n Aff	Aff Cues	Hi	Lo	
H	H	7	7	14
	L	10	11	21
L	H	15	13	28
	L	8	12	20
		40	43	83

Essay Test Scores
Women

n Aff				
H	H	18	18	36
	L	13	21	34
L	H	15	19	34
	L	25	16	41
		71	74	145

Table 4

The Interaction of n Affiliation, Teacher Affiliation Cues, and Performance on Final Course Examination.

(Second Study)

15-item Test

Men

n Aff	Aff Cues	Test Scores* H	L	Total
H	H	21	12	33
	L	19	12	31
L	H	32	43	75
	L	56	39	95
		128	106	234

Women

n Aff	Aff Cues	H	L	Total
H	H	46	36	82
	L	46	47	93
L	H	25	23	48
	L	31	22	53
		148	128	276

Interaction X^2 (all subjects)
Aff cues x n Aff x Knowledge= 4.25
df = 1
$p < .05$

Criteria Test Score

Men

n Aff	Aff Cues	H	L	Total
H	H	26	14	81
	L	19	19	78
L	H	54	46	59
	L	65	53	78

Women

n Aff	Aff Cues	H	L	Total
H	H	53	58	111
	L	62	64	126
L	H	30	31	61
	L	27	46	73
		172	199	361

* The test score distribution was split as close to the median as possible. Since a number of scores fill at the median, splits are not 50-50.

References

Atkinson, J.W. (Ed.) Motives in fantasy, action, and society. Princeton: Van Nostrand, 1958.

Cogan, M.L. The behavior of teachers and the productive behavior of their pupils: II. Trait analysis. J. exper. Educ., 1958, 27. 107-124.

McKeachie, W.J., Motivation, teaching methods, and college learning. In Jones, M.R. (Ed.) Nebraska Symposium on Motivation 1961. Lincoln, Nebraska. U. of Neb. Press, 1961.

Milholland, J.E. Measuring Cognitive Abilities. 1964. Sec. II-A of this report.

Pace, R.C., and Stern, G.C. A criterion study of college environment. Syracuse University, Psychological Research Center, 1958.

Ryans, D.G. Characteristics of teacher. Washington: American Council on Education, 1960.

Shipley, T.E. and Veroff, J.A. A projective measure of need affiliation. J. exp. Psychol. 1952, 43, 349-356.

Veroff, J., Atkinson, J.W. Feld, Sheila, and Gurin, G. The use of thematic apperception to assess motivation in a nationwide interview study. Psychol. Monogr., No. 499, 1960, 74, 12.

Student Achievement Motives, Achievement Cues, and Academic Achievement[1]

by

W.J. McKeachie, Robert L. Isaacson, John E. Milholland, and Yi-Guang Lin

When McClelland et. al. (1953) named their book and their measure of motivation "the achievement motive" they instigated one of the most frustrating yet tantalizing chases after a will-o'-the-wisp in the history of psychology. Obviously a measure of achievement motivation should predict achievement, and what index of achievement lies closer at hand than academic grades? The very terms "over-achievement" and "under-achievement" issue a mating call to measures of achievement motivation. Investigators hoped that here at last was a measure which would enable them to account for the considerable variance in college grades that was unrelated to intellectual measures.

McClelland and his colleagues ran the first leg of the resulting marathon, publishing results of studies at Wesleyan and Trinity in which correlations of .51 (N=30), .05 (N=40), and .32 (N=19) were found between grades and need for Achievement. Attempts to replicate these findings have been numerous and seldom published because of the generally negative results. Among positive findings were those of Morgan (1952, 1953); among negative results noted in published articles are studies by Atkinson (1950).

How can this seeming gap in construct validity be reconciled with other evidence supporting the validity of the measure? One plausible explanation is that academic achievement is over-determined (Atkinson, 1958, pp. 287, 601-603). By overdetermination, Atkinson means that when a number of motives may be satisfied by the same goal, a measure of any one of them is unlikely to be a good predictor. Because academic achievement may lead to satisfaction of needs for understanding, power, or affiliation, among others, it is unlikely to be predicted well by a measure of just one motive.

---

[1] The earlier data reported in this study were collected under a grant from the Fund for Advancement of Education.

The present research was designed to test another (not necessarily contradictory) possible explanation--that the achievement motive interacts with cues in the situation in influencing academic achievement. Because of these interactions, there are no uniform overall effects of need for achievement.

Hypotheses

Sex differences play an important role in the studies of achievement motive. Almost all successful studies have used men as subjects; attempts to use the measure with women have had only spotty success. In our studies, therefore, the data for men and women were treated separately and we anticipated more success with men.

Our basic hypothesis was thus that men high in achievement motivation would perform best in classes where the instructor set high achievement standards, encouraged competition, and where students perceived the work to be challenging. On the other hand, we hypothesized that men in the middle range of n Ach scores would do poorly in such sections. Our rationale for this was based on the suggestion made by McClelland and his colleagues that these men are characterized by a high fear of failure and that the anxiety engendered by a high level of achievement cues would interfere with their performance.

No prediction was made for students in the lowest third of the n Ach score distribution.

Procedure

To test our hypotheses we needed a sample of students differing in motivation and a sample of instructors differing in the cues emitted. We were primarily concerned with relative achievement within a class as a dependent variable, but there are some advantages to having several teachers trying to achieve the same goals, since this permits investigating comparative overall effects of differing methods. Moreover, there are clear advantages of generalization if conclusions can be based upon more than one course. Consequently, in our first experiment we chose three multi-section courses representing quite different types of content and objectives. They were second year French, freshman mathematics, and general psychology. The French and mathematics courses met four times a week; the psychology courses three times a week. Section sizes were typically 20 to 30. Thirty-one instructors participated, all of whom had had previous teaching experience. Except for four instructors in French, all instructors were men.

Within these courses all 825 students were invited to take a battery of tests designed to measure our independent variables. In psychology the time spent was credited as fulfilling a requirement that each student have first hand experience with research and almost every student participated. In French and mathematics we had less control over student participation but succeeded in testing some 80% of the men and 90% of the women.

In a second experiment only the introductory psychology course was involved. The 16 instructors in this experiment were individually responsible for two sections of up to 30 students each. Each section met four times a week. Four hundred six women and 348 men were enrolled in the course. About 95% of these students took the measures used in our study.

The measure of motivation was the TAT-type measure of McClelland et. al. (1953). The pictures used were those described by Veroff, Atkinson, Feld and Gurin (1960). Scoring was carried out by scorers whose reliability as measured by correlations with an expert ranged from .82 to .90. Because scoring extended over a three month period a later check on scorer reliability was made. At this time reliability of scores ranged from .81 to 1.00 (rho) for a 30-story sample.

High scores on this measure were taken as indicative as strong positive motivation for achievement. Scores in the middle third of the distribution were taken as indicative of fear of failure (Moulton 1958).

The measure of teacher achievement cues in the first study was the mean rating by a teacher's students on the following items which were included on a 12 item scale described by McKeachie (1961).

Instructor set very high standards for the students.
Members of the class competed to do well.
The course work presented a real challenge to me.

In our second study we used a single item from the 46-item scale described by Isaacson et. al. (1964). (None of the items used earlier had survived the criteria used for selection of items for this scale.)

He maintained definite standards of student performance.

The outcome measure used in both studies was course grade.

Results

We looked first for an overall relation between n Ach scores and grades, and as usual, found none. A summary of the data on this question from both studies is given in Table 1. None of the differences is statistically significant, but for the women the data are in the direction of a positive correlation between n Ach and grades.

Table 1

Relation Between n Ach Scores and Course Grades

Level of n Ach Score	Course Grade Men A or B	 Men C or D	 Women A or B	 Women C or D
High	62	78	138	109
Medium	55	62	98	103
Low	88	115	115	129

The presentation of results in the detail that would permit a test of our hypothesis appears in Tables 2 and 3. The data directly bearing on the hypothesis have been extracted and presented in Table 4. The only group for whom the pattern of performance is in accordance with the hypothesis is the Psychology group in the Second Study and here the trend is so slight as to be negligible. When all courses are combined, the results are in direct contradiction to the hypothesis, although the difference is not statistically significant.

The data for women in high Achievement Cue sections are more in accord with the hypothesis, although again the difference is not statistically significant. The table for all courses combined is presented below.

n Ach Score	Course Grades A or B	 C or D
High	52	43
Middle	29	34

Two *post hoc* hypotheses seemed worth checking. One was that students with low n Ach scores would need extrinsic sources of motivation, hence would perform better in high Achievement Cue sections. There was found to be a slight contrary trend for men, and an even more slight one for the women.

The second hypothesis to be looked into was that there would be a pervasive Achievement Cue effect. This also failed to appear.

Table 2

Interactions of Sex, n Ach, and Achievement Cues with Grades (First Study)

			Course Grades					
			Men			Women		
Course	n Ach	Ach cue	A or B	C or D	Total	A or B	C or D	Total
French	H	H	0	5	5	8	5	13
		M	1	2	3	14	7	21
		L	3	1	4	12	5	17
	M	H	0	4	4	7	2	9
		M	1	1	2	7	9	16
		L	0	0	0	8	5	13
	L	H	1	5	6	6	4	10
		M	2	3	5	9	12	21
		L	1	6	7	6	9	15
Mathematics	H	H	3	10	13	10	3	13
		M	3	3	6	2	0	2
		L	7	2	9	3	2	5
	M	H	4	8	12	6	7	13
		M	1	1	2	0	1	1
		L	6	3	9	3	4	7
	L	H	6	12	18	6	11	17
		M	3	7	10	1	2	3
		L	4	4	8	3	1	4
Psychology	H	H	1	6	7	9	8	17
		M	3	6	9	13	9	22
		L	1	6	7	5	5	10
	M	H	2	4	6	5	4	9
		M	3	3	6	8	13	21
		L	6	1	7	4	4	8
	L	H	2	2	4	5	3	8
		M	3	8	11	13	10	23
		L	4	4	8	1	8	9

Table 3

Interactions of Sex, n Achievement, and Achievement Cues with Grades in Psychology (Second Study)

			Course Grades		
	n Ach	Ach cues	A + B	C, D, and E	Total
Men	H	H	14	13	27
		M	13	10	23
		L	13	14	27
	M	H	13	14	27
		M	12	11	23
		L	7	12	19
	L	H	24	23	47
		M	20	21	41
		L	18	21	39
Women	H	H	25	27	52
		M	22	19	41
		L	15	19	34
	M	H	11	21	32
		M	18	15	33
		L	21	18	39
	L	H	24	29	53
		M	19	20	39
		L	22	20	42

Table 4

Course Grades of Men with High and Middle Range n Ach Scores Who Were in High Achievement Cue Sections

		Course Grades	
Course	n Ach Score	A or B	C or D
French	High	0	5
	Middle	0	4
Mathematics	High	3	10
	Middle	4	8
Psychology (First Study)	High	1	6
	Middle	2	4
Psychology (Second Study)	High	14	13
	Middle	13	14
All courses combined	High	18	34
	Middle	19	30

Discussion

The generally negative results reported here indicate that the refinements we introduced into our experimental design were not sufficient to bring forth the expected relations between achievement motivation and achievement. If any do in fact exist our measures must have been too crude to detect them or other uncontrolled variables (e.g. level of general scholastic ability) obscured them. We introduced only one new variable: level of Achievement Cues, and we must conclude that its interaction with n Ach is not sufficiently vigorous to produce any substantial effect on course grades.

References

Atkinson, J.W. Studies in projective measurement of achievement motivation. Unpublished doctoral dissertation. Univ. of Michigan, 1950. (Abstract in Univ. Microfilms, Vol I, *4*, Publication No. 1945.)

Atkinson, J.W. (ed.) *Motives in Fantasy, Action and Society*. Van Nostrand, Princeton, N.J., 1958.

Isaacson, R.L. et. al. The Dimensions of student evaluations of teaching, 1964. Section IV-A of this report.

McClelland, D.C. et. al. *The Achievement Motive*. Appleton-Century Crofts, New York, 1953.

McKeachie, W.J. Motivation teaching methods, and college learning. In Jones, M.R. (ed.) *Nebraska Symposium on Motivation* 1961. Lincoln, Neb. U. of Nebraska Press, 1961.

Milholland, J.E. Measuring Cognitive Abilities, 1964. Section II-A of this report.

Morgan, H.H. A psychometric comparison of achieving and nonachieving college students of high ability. *J. consult. Psychol.*, 1952, *16*, 292-298.

Morgan, H.H. Measuring achievement motivation with "Picture interpretations." *J. consult. Psychol.*, 1953, 17, 289-292.

Moulton, R.W. Notes for a projective measure of fear of failure. In Atkinson, J.W. *op. cit.* 1958.

Smith, C.P. and Feld, Sheila. How to learn the method of content analysis for *n*. Achievement, *n* Affiliation, and *n* Power. In Atkinson, J.W. *op. cit.* 1958.

Veroff, J., Atkinson, J.W. Feld, Sheila, and Gurin, G. The use of thematic apperception to assess motivation in a nation-wide interview study. *Psychol Monogr.*, No. 499, 1960, *74*, 2.

Student Power Motives, Discussion Teaching, and Academic Achievement[1]

W.J. McKeachie

This study is one of a series concerned with interactions between student motives and cues from instructors in relation to achievement in college classes. The theoretical context for these studies is described in McKeachie (1961).

Veroff (1957), who developed the scoring system for the power motive, describes the power motive as that "disposition directing behavior toward satisfactions contingent upon the control of the means of influencing another person." He found that students who scored high in power motivation enjoyed leadership and recognition and were rated by instructors as high in argumentativeness and in frequency of attempts to convince others of their point of view in the classroom. We thus hypothesized that the student high in power motivation would achieve well in classes characterized by a high proportion of student assertion.

Subjects

To test our hypotheses we needed a sample of students differing in motivation and a sample of instructors differing in the cues emitted.

In our first study, we chose three multi-section courses representing quite different types of content and objectives. They were second year French, freshman mathematics, and general psychology. Thirty-one instructors participated - all of the experienced teachers in these courses. Within these instructors' sections all students were invited to take a group of tests designed to measure our independent variables. Of the 825 students in the three courses, 609 gave us complete, useable responses on all tests. In our first sample there were 208 students in French, 188 in mathematics, and 213 in psychology. Two hundred fifty-four were men, 355 were women.

---

[1] The earlier data reported in this study were collected under a grant from the Fund for Advancement of Education.

The second study was carried out in an introductory psychology course. The sixteen instructors in this experiment were individually responsible for two sections of up to 30 students each. Each section met four times a week. Four hundred six women and 348 men were enrolled in the course. About 95% of these students took the measures used in our study.

Measures

Motives and Ability

Our measure of motives was the thematic apperception measure of achievement, affiliation, and power motivation described by Atkinson (1958). The pictures used were those used in the national survey of mental health by Veroff, Atkinson, Feld, and Gurin (1960). Separate pictures were used for men and for women.

As a measure of ability we used the scores on the American Council on Education Psychological Examination (1949 Form) which students had taken upon entering the University.

Classroom observers used a modification of Bales system of recording interactions (Mann, 1959). The system records each act in one of several categories. We combined some categories which were rare or unimportant in classroom observation. Rather than record each student's acts separately, we treated the classroom as a two-person group made up of a teacher and students. We had students' and teacher's responses classified into eight categories.

1. Positive reactions
2. Tension release
3. Suggestions
4. Orientation-opinion
5. Asking for orientation or opinion
6. Asking for support
7. Tension
8. Negative reactions

Observers were trained by Dr. Richard Mann. Comparisons between observers were made by plotting the number of acts placed in each category on binomial probability paper (Mosteller and Tukey, 1949). An observer's data were not used until his reliability with other observers consistently fell within the 5% boundary. Observers typically visited each class three times during the semester. In the first study this was sufficient to achieve a reliability in which variability between sections was significantly greater than variability between observers within a section at different times. (Medley and Mitzel, 1958) In the second study satisfactory reliability was not achieved.

Our index of student assertion was derived by dividing the number of student acts in Category 4 (Orientation-opinion) by the sum of student acts in Category 4 and Teacher acts in Category 5 (Asking for orientation and opinion). This seemed to us a reasonable measure of the degree to which students felt free to volunteer their own opinions other than in response to a question from the instructor.

Outcomes

We had two types of outcome measures. The most obvious one was the course grade.

A second group of outcome measures were achievement tests given to the psychology students. In the first study students had as part of their final examination 15 multiple-choice items and one essay question. All essay tests were scored blind by a single experienced instructor who was not involved in the experiment. In the second study the same 15-item objective test of knowledge was used and the Criteria test (Milholland, 1964) was also administered.

Results

In our first study men high in the power motive did get better grades in classes in which student volunteering was prevalent than in classes in which it was uncommon. The converse was true for men low in power motive (See Table 1). In our second study the results are much less marked although tending in the same direction.

The results for the first study when test scores were the outcome measure are shown in Table 2. On the objective examination both kinds of students averaged better in high assertion sections, and high _n_ Power students averaged better in both kinds of sections, so there was no interaction. On the essay examination the interaction appeared in the expected direction but at a low level of significance. In the second study the results tend in the expected direction in the objective test. (See Table 3)

There was no interaction effect for any of the women's data, so they are not reported here.

Table 1

The Interaction of n Power and Opportunity for Assertion with Grades for Men

First Study

Assertion	n Power	A + B	C + D	Total
High	H	25	30	55
	L	17	39	56
Low	H	12	27	39
	L	19	22	41
		73	118	191

Source	Interaction X^2	df	p
Assertion x n Power	0.0	1	NS
Assertion x grade	0.1	1	NS
n Power x grade	0.1	1	NS
Assertion x n Power x grade	3.9	1	.05

Second Study

Assertion	n Power	A + B	C + D	Total
H	H	29	29	58
	L	31	38	69
L	H	31	35	66
	L	35	34	69
		126	136	262

Source	Interaction X^2	df	p
Assertion x n Power	.272	1	.60
n Power x Grade	.008	1	.95
Assertion x Grade	.071	1	.80
n Power x Assertion x Grade	.508	1	.50

Table 2

The Interaction of n Power and Opportunity for Assertion as Related to Achievement Test Performance

(First Study)

Men

Objective Exam Score

	Student Assertion	
n Power	High	Low
H	$\overline{X}$ = 12.68 (N = 19)	$\overline{X}$ = 12.[illegible]3 (N = 19)
L	$\overline{X}$ = 12.32 (N = 19)	$\overline{X}$ = 10.73 (N = 11)

Essay Exam Score

	Student Assertion	
n Power	High	Low
H	$\overline{X}$ = 3.84 (N = 19)	$\overline{X}$ = 3.05 (N = 19)
L	$\overline{X}$ = 3.05 (N = 19)	$\overline{X}$ = 3.83 (N = 11)

n Power x Student Assertion: F = 3.36
df = 1 and 40
$p < .1$

Table 3

The Interaction of n Power and Opportunity for Assertion as Related to Achievement Test Performance

(Second Study)
Objective Exam

Men

n Power	Bales SASST	Hi	Lo	Total
H	H	31	20	51
	L	29	27	56
L	H	26	28	54
	L	34	25	59
		120	100	220

Source	Interaction X^2	df	p
n Power x BSASST	.000	1	.99
n Power x OB Test	.197	1	.70
BSASST x OB Test	.005	1	.98
n Power x BSASST x OB Test	1.888	1	.20

CPTS

n Power	Bales SASST	Hi	Lo	Total
H	H	29	33	62
	L	35	36	71
L	H	36	40	76
	L	38	35	73
		138	144	282

Source	Interaction X^2	df	p
n Power x BSASST	.543	1	.40
n Power x CPTS	.067	1	.80
BSASST x CPTS	.364	1	.60
n Power x BSASST x CPTS	.034	1	.60

Discussion

With a replication that is only weakly supported, a lengthy discussion appears unwarranted. Nevertheless we have somewhat greater confidence in our hypothesis than when we began these studies. It appears that men high in the power motive are more highly motivated in sections where there is a good deal of opportunity to volunteer one's own opinions. Moreover the results are not entirely due to their making a more favorable impression upon the instructor, for they are found in test scores as well as grades.

We were disappointed by the lack of positive findings for women. Perhaps power motivation is not emphasized as an important aspect of socialization for girls, or perhaps situations permitting assertion are not those eliciting power motivation in women.

Summary

A permissive teaching style, in which student volunteering of opinion was prevalent, affected achievement of students differentially in one study. The effect was favorable for men high in power motive. The trend was replicated in a second study but not at a very encouraging level of significance.

References

Atkinson, J.W. (Ed.) Motives in Fantasy, Action and Society. Van Nostrand, Princeton, N.J. 1958.

McKeachie, W.J. Motivation, Teaching methods, and college learning. In Jones, M.R. (Ed.) Nebraska Symposium on Motivation 1961. Lincoln, Neb. U. of Neb. Press, 1961.

Mann, Richard D. The relation between personality characteristics and individual performance in small groups. Doctoral dissertation, Univ. of Michigan. 1959

Medley, D.M. and Mitzel, H.E. Application of analysis of variance to the estimation of the reliability of observations of teachers' classroom behavior. J. exper. Educ., 1958, 27, 23-35.

Mosteller, F. and Tukey, J. The use and usefulness of binomial probability paper. J. Amer. Statistical Association. 1949, 44, 174-212.

Veroff, J. Atkinson, J.W., Feld, Sheila, and Gurin, G. The use of thematic apperception to assess motivation in a nationwide interview study. Psychol. Monogr., No 499, 1960, 74, 2.

Veroff, J. Development and validation of a projective measure of power motivation. J. abnorm. soc. Psychol., 1957, 54, 1-8.

Veroff, J., Wilcox, Sue, and Atkinson, J.W. The achievement motive in high school and college age women. J. abnorm. soc. Psychol. 1953, 48, 108-119.

VI-B. INTERACTIONS OF TEACHER CHARACTERISTICS AND STUDENTS' SCORES ON THE OPINION, ATTITUDE AND INTEREST SURVEY WITH OUTCOME MEASURES

John E. Milholland

The Opinion, Attitude and Interest Survey (Fricke, 1955; 1962) was developed to provide personality measures that correlate with performance in college. Since 1957 it has been administered to freshmen at The University of Michigan and scores on it were available on students in our sample who came as freshmen. Five of its 14 scores were selected for study in this research: Achiever Personality (AP), Intellectual Quality (IQ), Social Adjustment (SA), Creative Personality (CP), and Social Science Interest (SSI). All scales were empirically derived, so have no theoretical rationale.

Achiever Personality usually correlates between .30 and .40 with college grades, but negligibly with tests of scholastic aptitude. Intellectual Quality correlates with interests in intellectual activities, and with aptitude tests; Social Adjustment and Creative Personality were validated against ratings by counselors, residence hall advisers and teachers; and the Social Science Interest Scale differentiates students who major in social sciences from those who do not.

For each OAIS scale two subgroups were used: those in the upper 30% and those in the lower 30% of our score distribution. Each subgroup was further divided for each teacher attribute: Achievement Cues, Skill, and Culture, according to whether the student was in a class whose teacher was above or below the median. (The method of scoring teachers is described in Section VII of this report.) For each combination of OAIS scale and teacher attribute, then, there were four groups to be compared on our three outcome measures: course grade 15-Item Test Score, and Criteria Test Score. The number of cases in each group is shown in Table 1. The students were in the introductory psychology course in the Spring Semester, 1960-61.

The relations between OAIS scores, teacher characteristics, and performance on the Criteria Test, the 15-item content test, and course grades are indicated in Table 2. The entries in the cells describe the attribute of the teacher that made for better student performance in our sample. Thus it appears that men students in both the high and low 30% of the Achiever Personality distribution do better, on both tests and on grades, with teachers who are high in Achievement Cues. Women, however, exhibit an interaction effect. On two of the three measures high AP women do better with low Achievement Cue teachers, whereas low AP women do better with high Achievement Cue teachers. This interaction would provide a basis for assigning women to teachers on the basis of student AP and teacher Achievement Cues. Similarly, both men and women could be assigned to teachers on the basis of student AP and teacher

Culture, but the method of assignment is different for the two sexes. For women, the relations of AP and Skill to grade and to tests are different, but there is some evidence that low AP men should be assigned to high Skill teachers. A description of how the relations of the five OAIS variables to teacher characteristics and outcome measures were used in assigning students to teachers in the Spring Semester of 1963 appears in Section VII of this report.

Reference

Fricke, B. G. Opinion, Attitude and Interest Survey. Ann Arbor: Evaluation and Examinations Division, The University of Michigan, 1962.

TABLE 1

NUMBER OF STUDENTS IN EACH SUBSAMPLE

OAIS Scale	Student Score	Teacher Attribute											
		Achievement Cues				Skill				Culture			
		Men		Women		Men		Women		Men		Women	
		Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
AP	High	21	33	38	44	29	29	23	19	40	21	30	30
	Low	18	24	33	43	34	17	34	30	27	30	37	40
IQ	High	15	28	38	35	27	25	29	23	22	27	33	38
	Low	23	30	29	45	36	24	31	26	26	26	17	34
SA	High	15	24	34	43	31	20	29	22	28	19	28	33
	Low	22	30	36	42	27	28	29	34	19	27	38	35
CP	High	17	33	28	39	26	26	22	26	22	28	35	37
	Low	23	25	34	38	29	21	32	22	26	20	36	29
SSI	High	19	33	36	44	29	25	25	23	23	30	32	36
	Low	22	23	30	42	27	21	33	32	22	24	38	34

TABLE 2

TEACHER CHARACTERISTICS (H = ABOVE MEDIAN; L = BELOW MEDIAN) FAVORING BETTER PERFORMANCE BY STUDENTS HIGH (UPPER 30%) OR LOW (LOWER 30%) ON EACH OAIS SCALE

OAIS Scale	Student Score	Performance Measure	Teacher Characteristics					
			Ach. Cues		Skill*		Culture	
			Men	Women	Men	Women	Men	Women
AP	High	Criteria Test	H	L	L	H	H	L
		15-Item Test	H	H			H	H
		Grade	H	L	H	L	H	L
	Low	Criteria Test	H	H	H	H	L	H
		15-Item Test	H	L			H	H
		Grade	H	H	H	L	L	H
CP	High	Criteria Test	L	L	L	H	L	H
		15-Item Test	H	L			L	L
		Grade	L	H	H	H	H	H
	Low	Criteria Test	H	H	H	H	H	H
		15-Item Test	H	H			H	H
		Grade	H	L	H	L	L	L
IQ	High	Criteria Test	L	H	L	L	L	H
		15-Item Test	L	H			H	L
		Grade	L	H	L	H	H	H
	Low	Criteria Test	H	H	L	H	L	L
		15-Item Test	L	H			L	H
		Grade	L	H	H	L	L	L
SA	High	Criteria Test	H	L	L	H	H	H
		15-Item Test	H	L			H	L
		Grade	H	L	H	H	L	H
	Low	Criteria Test	L	H	L	H	L	H
		15-Item Test	L	H			L	H
		Grade	H	H	H	L	L	H
SSI	High	Criteria Test	L	H	L	H	L	H
		15-Item Test	L	**			L	H
		Grade	L	H	H	H	H	H
	Low	Criteria Test	H	L	H	L	L	H
		15-Item Test	H	H			H	L
		Grade	H	L	H	L	L	L

*The analysis for teacher skill with the 15-item test was not made, through oversight.

**Equal

STUDENT RATINGS OF TEACHERS AND STUDENT PERFORMANCE AS A FUNCTION OF STUDENT AND TEACHER PERSONALITY FACTORS

James W. Baerwaldt, John E. Milholland, and W. J. McKeachie

This section reports the results of a study designed to discover possible interactions of student and teacher personality variables affecting student performance and satisfaction in introductory psychology.

Method and Results

Subjects were sixteen teaching fellows and about 570 students in the introductory psychology course in the spring semester of 1961.

The teachers took Forms A and B of the IPAT Sixteen Personality Factors Questionnaire (16 PF) and the Peer Group rating form described in Isaacson et al. (1963). The students took Form A of the 16 PF and, at the end of the semester, filled out a 46 item rating of their instructors and took the Criteria Test. The American Council on Education Psychological Examination (ACE) had been administered to the students when they entered the University. These scores were used to partial ability out of the Criteria Test scores. The resulting scores were called the residual Criteria Test scores.

Responses on the Student Rating Form had been factor analyzed and five stable factors identified (Isaacson, McKeachie, Milholland, 1963; Section IV-B of this report).

CORRELATIONS OF STUDENT RATING FACTOR SCORES AND TEST SCORES

The items with high loadings on a factor were used to obtain a teacher's score for that factor. The correlations of the teacher mean factor scores with each other and with the student mean residual Criteria Test Scores for the sixteen teaching fellows are shown in Table 1. No evaluation of statistical significance was made because of probable lack of normality in the distributions and the fact the correlations were between mean scores rather than individual scores. The discussion below will deal with the correlations as given.

Average teacher ratings given by men and by women students agreed fairly well with respect to Skill and Structure scores, moderately well on Overload and on Student-Teacher Rapport, and poorly on Feedback and Group Interaction. This pattern seems reasonable in showing higher agreement on Skill and Struc-

TABLE 1

INTERCORRELATIONS, ACROSS 16 TEACHERS OF INTRODUCTORY PSYCHOLOGY, OF MEAN FACTOR SCORES DERIVED FROM STUDENT RATING ITEMS, AND CORRELATIONS OF THESE SCORES WITH MEAN RESIDUAL CRITERIA TEST SCORES*

Factors	Factors						Residual Criteria Test Score	
	1	2	3	4	5	6	Men	Women
1. Skill	.70	-.10	.34	.67	.72	.65	.39	.28
2. Overload	.13	.58	.40	-.26	.02	-.14	.14	.45
3. Structure	.10	-.47	.69	-.06	.32	.36	.25	.01
4. Feedback**	.74	-.08	.04	.32	.39	.35	.43	.10
5. Group Interaction	.63	-.03	-.12	.44	.30	.68	.05	.13
6. Student-Teacher Rapport	.78	-.08	.03	.78	.50	.46	.64	.31

*Correlations for men students are above the diagonal; those for women are below. Correlations in the diagonal are between the mean factor scores based on men's ratings and on women's ratings.

**This factor does not correspond precisely with the Feedback factor described in Section IV-A of this report, but includes the items defining it.

ture, which are fairly objective, but one might expect Group Interaction and Feedback to be more objective than Overload and Student-Teacher Rapport.

The method employed to factor analyze the student ratings yielded factors that were orthogonal, but of course the factor scores, derived from highloading items, need not be. Some of the correlations in the table, however, seem to be excessively high. There seems to be some kind of averaging effect in operation, since the correlations across students of factor scores are considerably lower (Skill vs. Rapport, for example, correlated .37 for men and .41 for women). On the other hand, most of the high correlations involve the Skill score, which has the largest number of items in it and may represent something approaching a general factor score. The exceptions to this are the correlation of Group Interaction with Student-Teacher Rapport (.68) for the men and the correlation of Feedback with Rapport (.78) for the women.

The Skill and Student-Teacher Rapport factor scores are the only ones consistently correlated with residual Criteria Test scores, while Group Interaction consistently shows no appreciable relation. The other three factor scores are more highly correlated for one sex than for the other. The nature of the data does not warrant a serious effort to interpret these discrepancies. There is a temptation, however, to say that if you want women to do well you should give them heavy work loads, and that men do better in more highly structured courses with feedback, whereas these conditions make no difference to women.

RESULTS OF ANALYSES OF VARIANCE

Four analyses of variance for men students and four for women students were performed using student and teacher personality measures as independent variables and student rating scores and residual Criteria Test scores as dependent variables. A listing of the variables used in each analysis is given in Table 2. The statistically significant results are discussed in the following paragraphs.

STUDENT CHARACTERISTICS RELATED TO ACHIEVEMENT AND RATING OF INSTRUCTOR

The only student characteristic which standing alone was predictive of the scores on the Criteria test was introversion, which was positively related to residual Criteria test scores for women (Analysis I).

Certain student characteristics were related to student ratings of teachers. For example, in Analysis III women's scores on Q_1 (Radicalism) were negatively related to their ratings of teachers on Overload and positively related to their ratings of teachers on Group Interaction.

TABLE 2

VARIABLES USED IN THE ANALYSES OF VARIANCE

Measures Used	Analysis Number			
	I	II	III	IV
Independent Variables				
Cattell 16 PF				
First Order Factors				
1. E. Dominance: Teacher		X		
2. Student		X		
3. N. Sophistication: Student				X
4. Q_1 Radicalism: Teacher			X	
5. Student			X	
Second Order Factors				
6. Anxiety: Teacher	X			
7. Student	X			
8. Introversion: Teacher	X			
9. Student	X			
Peer Group Ratings				
10. Scale 5. Cultural Attainment: Teacher				X
Dependent Variables				
Student Rating Form				
1. Item 1. Overall Teaching Ability	X			
2. Item 2. Overall Value of Course	X			
Factor Scores				
3. Skill		X	X	X
4. Overload		X	X	X
5. Structure		X	X	X
6. Feedback		X	X	X
7. Group Interaction		X	X	X
Criteria Test				
8. Residual Score	X	X	X	X

Similarly, women's scores on Dominance were related to their ratings of instructors on Feedback (Analysis II). Women in the middle third of the distribution of Dominance scores perceived the most Feedback, high dominant women the least.

INSTRUCTOR CHARACTERISTICS RELATED TO STUDENT RATINGS

Personality characteristics of instructors were also related to student ratings of them. A correlational analysis of these results has already been reported (Isaacson, McKeachie, and Milholland, 1963). The analyses of variance used in the present study did not change our conception of those results, but are presented for the sake of completeness.

Teachers high in Dominance tended to be given the highest Skill scores by both men and women (Analysis II). The high Dominance teachers were also perceived by men as providing a high degree of Feedback and Structure. Teachers who were above the median on Q_1 (Radicalism) were rated high on Skill and on Feedback (Analysis III). Male students also rated these instructors high on Group Interaction. Cultured teachers (as identified by ratings of peers) were rated high on Skill, Group Interaction, and Feedback, but Culture was negatively related to the ratings on Structure (Analysis IV).

Finally, women rated teachers who were high on both anxiety and introversion as being more able than teachers high on anxiety, low in introversion (Analysis I).

INTERACTIONS BETWEEN STUDENT PERSONALITY AND TEACHER PERSONALITY

These analyses had been undertaken with the purpose of identifying interactions between personality characteristics of teachers and personality characteristics of students. Would extraverted students, for example, do best with an extraverted teacher? Would a dominant teacher be most successful with students low in dominance? The results emerging from the analyses were not startling but nevertheless interesting.

One surprising finding, for example was that men high in Sophistication (N) did best on the Criteria Test in classes taught by teachers rated low on Culture. "Cultured" teachers were most effective with men scoring in the lower two thirds of the distribution of Sophistication scores (Analysis IV).

Men varying in Sophistication did not rate instructors differing in Culture as being different in Skill, but ratings of Feedback were affected. For high-Sophistication men ratings of Feedback were positively related to teacher Culture; middle-Sophistication men rated low-Culture men as being moderate in Feedback and mid-Culture teachers low in Feedback; low-Sophistication men rated the mid-Culture teachers highest in Feedback. In short the only agreement was that high-Culture teachers give more feedback than low-Culture teachers.

One other interaction effect emerged in student ratings. Women scoring high on Radicalism (Q_1) rated their high-Radicalism instructor high on the

"Overload" factor; mid-Radicalism women rated these instructors low on Overload.

Discussion

The pattern of significant results is spotty, but some of the trends appear to be reasonable. For example, it seems sensible to think that the more dominant teachers would be more likely to be rated higher on Skill and on Structure, and possibly on Feedback. It also seems not surprising that teachers rated high on Culture (by fellow teachers) would be rated low on Structure by their students. Explanations for the other significant results do not come readily to hand.

The residual Criteria Test Score appeared in all the analyses of variance, yet showed only two instances of relations with teacher or student personalities.

In the correlation study, however, it is gratifying to note that the ratings on skill did relate positively to Criteria Test performance, and that good Student-Teacher Rapport seems to be beneficial. It may be that student performance in the classroom, at the surface level at least, is so highly dependent on intellectual ability and orientation that measures of other variables will have to be extremely delicate and precise, and multiple controls imposed, in order for their effects to become apparent. Our experiences in this research have given us some leads for better designs and useful variables to devote attention to.

Reference

Isaacson, R. L., McKeachie, W. J., and Milholland, J. E., Correlation of teacher personality variables and student ratings. J. Educ. Psychol. 1963, 54: 110-117.

VII. THE MATCHING OF STUDENTS AND TEACHERS SO AS TO ENHANCE STUDENT ACHIEVEMENT

John E. Milholland

One of the principal foci of the entire research project has been the question of whether characteristics of students and teachers could be located which would "mesh" so that it could be said that certain kinds of students would perform best with certain kinds of teachers. The attributes which make for good college classroom performance in general have been well studied in the past, and the list of them is fairly short. Our own study of main effects (Section V-A of this report) has confirmed the importance of intellect, as defined by tests of scholastic aptitude, and the relatively minor roles played by other attributes, which are generally allocated to the personality domain, when a group of students is studied as a whole. It remained to be seen whether there were interactions present which would allow the latter kind of variables to exhibit more effect. The measures we had chosen were among those which, in one setting or another, had been related significantly to academic performance. We planned to use them to place students with teachers with whom they could be expected to do well.

The design for this phase of the project called for an assessment of both students and teachers during the 1960-61 academic year and the study of the relationships of the assessment measures, singly and in combination, to various measures of outcomes in the introductory psychology course. If the results showed promise they were to be used to develop descriptions of teachers and students that would enable us to assign students to classes in such a way as to favor their good performance. An experiment of this nature was scheduled for the academic year 1962-63. This section of our report describes that experiment.

It was fortunate for our purposes that The University of Michigan was moving into a system of pre-registration in the year that we were doing this research. Efforts were made to have as many students as possible actually complete their enrollment in their spring semester courses during the fall semester. Pre-registration was voluntary (it has since been made compulsory) but many students saw it as a chance to procure a preferred schedule of classes. About half of the approximately 1000 students who eventually enrolled for introductory psychology had pre-registered and been given our battery of assessment measures.

The assessment measures came from two sources. The freshman entrance testing program provided scores on the College Entrance Examination Board Scholastic Aptitude Test (SAT) and on the Opinion, Attitude and Interest Survey (OAIS) and we administered a set of TAT pictures, the N (sophistication) Scale of the Cattell 16PF, the Alpert-Haber Achievement Anxiety Test, the study habits inventory, and two attitude questionnaires in pre-testing sessions on December 12, 1962, January 15, 1963, and February 2, 1963.

Although we had originally planned to use the TAT scores in our matching, the time required to score the protocols precluded it. The measures we did use, and the patterns for matching are given in Table 1. It shows, for example, that a high SAT student should be placed with a high Achievement Cue teacher, but a low SAT student with a low Achievement Cue teacher.

The scoring procedures were set up after considerable study and discussion of the fall semester results. Statistically significant and consistent differences were sparse, making it necessary to introduce some subjectivity into decisions about scoring. We nevertheless decided to undertake the matching experiment, using some rather tenuous relationships, because it would help us decide whether effects were going to be powerful enough to be easily discernible or so delicate that more subtle methods would be called for. In addition, the Spring, 1963, classes would provide an opportunity to replicate much of the 1961 research, perhaps confirming what we observed then or furnishing new leads for exploration.

The measures for teachers had come from the fall semester student ratings on the items indicated.

Achievement Cues: Ratings of A (this almost always occurred) or
B (this often occurred)

Item 8: He maintained definite standards of student performance.
Item 31: He told students when they had done a particularly good job.
Item 48: He criticized poor work.

Skill: Ratings of A (almost always) or B (often)

Item 1: How would you rate your instructor in general (all-around) teaching ability? (A: an outstanding and stimulating instructor); (B: a very good instructor).
Item 6: Class time was well spent.
Item 7: The instructor was skillful in observing student reactions.
Item 22: He put his material across in an interesting way.
Item 44: He stimulated the intellectual curiosity of his students.
Item 45: He changed his approach to meet new situations.
Item 49: He explained clearly and his explanations were to the point.

Culture: Ratings of A (almost always) scored 5; B (often) 4; C (occasionally) 3; D (seldom) 2; E (never) 1.

Item 21: He exhibited a high degree of cultural attainment.

TABLE 1

BASIS FOR MATCHING STUDENTS AND TEACHERS

SPRING SEMESTER, 1962-63

Student Variable	Student Status	Teacher Characteristic Favoring Good Student Performance		
		Achievement Cues	Skill	Culture
Men Students				
Scholastic Aptitude	high	high		
	low	low		
Opinion, Attitude and Interest Survey Scales				
1. Achiever Personality	high		low	high
	low		high	low
2. Creative Personality	high	low	low	low
	low	high	high	high
3. Intellectual Quality	high		low	high
	low		high	low
4. Social Adjustment	high	high		high
	low	low		low
5. Social Science Interest	high	low	low	
	low	high	high	
Cattell 16PF Factor N	high			low
	low			high
Women Students				
Opinion, Attitude and Interest Survey Scales				
1. Achiever Personality	high	low		low
	low	high		high
2. Creative Personality	high	low	high	
	low	high	low	
3. Intellectual Quality	high			high
	low			low
4. Social Adjustment	high	low	high	
	low	high	low	
5. Social Science Interest	high	low	high	high
	low	high	low	low
Cattell 16PF Factor N	high			low
	low			high
Study Habits	high	low		
	low	high		

For Achievement Cues and Skill, teachers were scored one point for each A or B rating they had on the designated items; on the Culture item they were given scores as indicated. The mean rating for each instructor on each attribute was computed and the group divided at the median of the means into a high and low subgroup. The median scores were 1.37 for Achievement Cues, 4.43 for Skill, and 3.81 for Culture.

Each student was given a score to classify him as being suited best to a teacher who was high or low on each of the three teacher attributes. Distributions were first made of students' scores on the various pre-test measures and each student located in the upper or lower half of each distribution. He was then given a score of 1 or 0 for each teacher attribute, according to whether or not he was in the half of the distribution that represented compatibility with teachers who were scored high on that attribute. The scoring patterns are shown in Table 2, and the 1's and 0's there correspond with the "highs" and "lows" in Table 1. The score distributions used for the dichotomizations were those obtained from the 125 men and 241 women who had been tested on December 12 and January 15.

The weights assigned a student for one of the teacher attributes were summed to give him a score for that attribute. The scores for all students (of the same sex) were obtained and each student designated "high" or "low" in accordance with his position above or below the median. The scores used for this purpose are also shown in Table 2. These high-low allocations for the three attributes formed the basis for determining the matching between student and teacher.

In Psychology 101 two sections had been scheduled at each class hour, so that there would be an opportunity for shifting students to teachers they matched better. Restrictions on perfect matching were quite severe, of course, since there were eight possible combinations of allocations on the three teacher attributes and only two teachers available at a particular class hour. Shifting was done whenever the degree of correspondence between teacher and student could be improved.

In Psychology 100 there was no provision for transfer of students, since there were not enough students enrolled in it to justify having two sections every hour.

By the end of the semester the number of students for whom complete data (including course grades and Criteria Test Scores) were available had dwindled to 333: 156 men and 177 women. Their distribution with respect to their matching with their teachers is given in Table 3.

The most obvious contrast to make is between the performance of students who matched their teachers on all three categories (those in the last row of Table 3) and those (in the first row of the table) who did not match at all. These contrasts are made in Table 4. No difference is statistically signif-

TABLE 2

CUTTING SCORES AND SCORING WEIGHTS FOR STUDENT MEASURES IN RELATION TO EACH TEACHER ATTRIBUTE

Student Pre-test Measure	Student Position	Score Required	Weights for Teacher Attributes		
			Ach. Cues	Skill	Culture
Men					
Scholastic Aptitude Test V+M	high		1		
	low	below 1200	0		
Opinion, Attitude and Interest Survey Scale					
Achiever Personality	high			0	1
	low	below 47		1	0
Creative Personality	high		0	0	0
	low	below 55	1	1	1
Intellectual Quality	high			0	1
	low	below 66		1	0
Social Adjustment	high		1		1
	low	below 39	0		0
Social Science Interest	high		0	0	
	low	below 40	1	1	
Cattell 16PF: Sophistication (N)	high				0
	low	below 11			1
Score for Allocation to High Group			3,4	2,3,4	3,4,5
Women					
Opinion, Attitude and Interest Survey Scale					
Achiever Personality	high		0		0
	low	below 58	1		1
Creative Personality	high		0	1	
	low	below 36	1	0	
Intellectual Quality	high				1
	low	below 58			0
Social Adjustment	high		0	1	
	low	below 52	1	0	
Social Science Interest	high		0	1	1
	low	below 40	1	0	0
Cattell 16PF: Sophistication (N)	high				0
	low	below 9			1
Study Habits	high		0		
	low	below 20	1		
Score for Allocation to High Group			3,4,5	2,3	2,3,4

TABLE 3

NUMBER OF STUDENTS EXHIBITING EACH KIND OF MATCHING WITH THEIR TEACHERS SPRING SEMESTER, 1963 (x INDICATES MATCHING)

Matched Categories			No. of Students		Total
Achievement Cues	Skill	Culture	Men	Women	
			13	18	31
x			27	17	44
	x		11	14	25
		x	20	14	34
x	x		12	15	27
x		x	22	28	50
	x	x	29	48	77
x	x	x	22	23	45
Total			156	177	333

TABLE 4

PERFORMANCE OF STUDENTS IN INTRODUCTORY PSYCHOLOGY, WHO MATCHED THEIR TEACHERS ON THREE TEACHER ATTRIBUTES, AND ON NONE SPRING SEMESTER, 1963

Performance Measure	Men			Women		
	Number of Attributes Matched			Number of Attributes Matched		
	Three		None	Three		None
Criteria Test						
N	22		13	23		18
Mean	40.0		37.2	38.0		37.3
SD	7.02		6.27	7.25		5.72
t		1.15			.33	
Course Grade						
N	22		13	23		18
Mean	2.8		2.9	2.3		2.6
SD	.71		.83	.91		.95
t		-.37			-1.00	

icant, but for both sexes the matching seemed to have a positive effect on Criteria Test Scores and a negative effect on grades.

The relation between the mere number of attributes on which students and teachers were matched and student performance is shown in Table 5. The S in the next to last row of each subtable is the numerator used for the adaptation of Kendall's tau-coefficient to ordered contingency tables. It is simply the net number of individual pair comparisons in the direction indicated by the sign; positive indicating a higher performance score going with a closer match. The denominator used to obtain the tau-coefficients shown is an estimate of the total number of pair comparisons that could be made, taking into account the fact that comparisons within the same rows or columns of the table are eliminated. The four tau-coefficients are extremely low, obviously not statistically significant, but again the positive trend for test scores and the negative one for grades appear. The configurations of the data in the body of the table make it clear that this attempt to improve performance by matching students to teachers did not have very beneficial effects, as measured by grades or Criteria Test Scores. This result does not, however, seem to us to point unequivocally toward the abandonment of our position that a particular aspect of a teaching situation may affect different students differently. We think perhaps the idea did not have a fair trial and can cite several features of our procedure that may have operated against us.

It may be that the interaction effects are weaker and more delicately balanced than we made provision for. Our method of matching was crude, using rough dichotomizations and simply summing the 1-0 scores to get a total. Further, we used relative, rather than absolute, cutting scores. If there are actually critical points in the continua of the teacher or student attributes we measured, they might well have escaped us.

Our set of variables on which we based our matching was rather restricted, even though some selection had been made from a larger pool. One substantial omission was the scores derived from the TAT protocols, although at the time we did our matching we would not have known how to use them effectively.

Some indication that more refined analyses, perhaps with other variables included, would be useful, comes from a study of a portion of our data carried out by Margaret Schaefer. She thought that perhaps the students who preregistered early and also appeared for pre-testing the first time they were asked, without having to be followed up, would be a homogeneous group of particular interest. The results of her study appear in Table 6. In this case both test scores and grades favor the students who were better matched, with the difference in grades coming closer to statistical significance. Miss Schaefer obtained data on several other subgroups but, with one exception, all did more poorly than the group of assigned students. The exception is mystifying—a small group (18 for grades; 16 for Criteria Test) who outperformed every other group, making an average grade of 2.89 and a Criteria Test Score of 40.38. The thing they had in common was that they had neglected to record

TABLE 5

CRITERIA TEST SCORES AND COURSE GRADES IN RELATION TO EXTENT TO WHICH STUDENTS AND TEACHERS WERE MATCHED

Number of Attributes Matched	Criteria Test Score 10 -19	20 -29	30 -39	40 -49	50 -59	Total	Course Grade E	D	C	B	A	Total
						Men						
3		2	6	12	1	22			8	10	4	22
2		7	33	21	2	63	6	8	25	21	9	63
1	2	9	17	27	3	58		3	21	20	14	58
0		2	6	5		13			5	4	4	13
Total	2	20	62	66	6	156		11	59	55	31	156
S						202						- 582
Tau						.025						-.070
						Women						
3		1	12	10		23	1	1	13	5	3	23
2	1	11	39	38	2	91		6	26	40	19	91
1		3	22	18	2	45	2	1	17	15	10	45
0		2	10	6		18		1	10	2	5	18
Total	1	17	83	73	4	177	3	9	66	62	37	177
S						266						- 293
Tau						.027						-.028

TABLE 6

PERFORMANCES OF STUDENTS TESTED IN DECEMBER AND JANUARY IN RELATION TO WHETHER OR NOT THEY HAD BEEN ASSIGNED* TO THEIR TEACHERS

	Performance Measure					
	Criteria Test			Grade		
	Assigned		Not Assigned	Assigned		Not Assigned
N	132		214	137		243
Mean	38.79		37.76	2.72		2.53
SD	6.82		6.79	.93		.94
t		1.36			1.90	

*An assigned student was one who matched his own instructor better than the other one available at the same hour. All Psychology 100 students were considered as unassigned.

their testing date on their answer sheet, and by the time we discovered it their papers had been mixed in with others.

An uncontrolled aspect of the situation, which could loom large, is the possibility that the teachers may have changed between the time we measured them (in the fall) and the time we measured their students (in the spring). For many of the teachers, the fall semester's work was their first teaching experience. Two of the teacher attributes in particular, Achievement Cues and Skill, might have undergone change from fall to spring.

The data we have on hand, plus some analyses we have already made of data previously collected, will permit us to make a more intensive and thorough study of the effects of correspondence of teacher and student than we are able to present at this time. We have every intention of carrying through such a study.